Biosynthesis and NMR Spectroscopic Studies of Large C-Terminal Fragments of Human Y4 Receptor

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Harsha Vardhan Kocherla

aus

Indien

Promotionskomitee

Prof. Dr. Oliver Zerbe (Vorsitz und Leitung der Dissertation)

Prof. Dr. Raimund Dutzler

Prof. Dr. Roland Riek

Zürich, 2014

Table of Contents

Chapter III: Expression of C-terminal NPY4R fragments in inclusion bodies (IBs) and their purification

Chapter IV: Biophysical characterization of the C-terminal NPY4R fragments by NMR

Chapter V: Biophysical characterization of the C-terminal NPY4R fragments by NMR

Summary

G protein-coupled receptors (GPCRs) belong to a group of alpha-helical transmembrane proteins (TMPs) that are constantly involved in triggering vital biochemical responses. These receptors exist in all eukaryotic organisms, very diverse in nature and especially in humans, almost 800 varieties of genes were identified that code for these TMPs. GPCRs totally constitute only 1% of the total cellular proteins, but they mediate many vital functions of immense pharmacological significance. Almost 50% of the drugs on the market and in research target these receptors with therapeutic implication against many diseases. The major role of GPCRs is to broadcast a signal across the cellular membrane. GPCRs function by receiving an extracellular signal from ions, photons, small organic molecules, peptides, and entire proteins (for example hormones) resulting in conformational changes that affects the interacting G-proteins inside the cells. After receiving the signal G-proteins dissociate from the GPCRs, split into two subunits that trigger a cascade of downstream signaling events. It is important to note that the fundamental event triggered by ligand binding is a conformational change, which is exploited by many drugs. The relationship between structure and biological action is of great interest to tackle existing diseases or disorders because it allows designing drugs that will interfere with GPCR signaling.

In humans, approximately 15% of all genes code for GPCRs, but the function of most of them has not been discovered yet. For those the function is known, structural information is limited. Even with the present advanced technologies, characterizing these receptors at atomic-resolution is extremely challenging because of some of their properties such as their hydrophobic nature, low natural abundance, high flexibility and the lack of finding an ideal membrane mimic for them.

In my first chapter, I introduce the properties of these receptors and subsequently elaborate on the recent developments in trying to overcome low expression levels by using heterologous strains. I conclude the chapter by discussing the possibility to use different membrane mimetics that are suitable for reconstituting TMPs for biophysical analysis.

In chapter 2 and 3, I present a divide-and-conquer approach towards solution structures of G-protein coupled receptors. Therein the human Y4 receptor is dissected into 2-3 transmembrane helix fragments, which are individually studied by solution NMR. I systematically compare various aspects of biosynthetic routes for expression of the fragments in *E. coli* and discuss purification strategies. In particular, I compare the production of the TM fragments in inclusion bodies using the insoluble ΔTrp leader sequence, with the membrane-directed expression using Mistic as the fusion partner. Moreover, I have developed methods for enzymatic cleavage to liberate the protein of interest from the fusion partner. The direct expression of 2 and 3-TM fragments into inclusion bodies are successful for the TMP discussed in the third chapter.

Biophysical evaluation of membrane proteins using solution NMR spectroscopy technique has recently advanced and in the fourth chapter, I describe the expression, purification and subsequent NMR-spectroscopic characterization of all the C-terminal NPY4R fragments. I also discuss the potential in transferring chemical shifts onto larger fragments based on common NOESY correlations. I conclude the chapter discussing possible explanations for the missing or unassigned peaks.

Finally, in chapter 5 I describe one particular membrane environment developed by Sligar and colleagues that mimics native membranes very closely. These membrane mimetic entities are called nanodiscs, and are made up of discoidal phospholipid bilayers (~ 10-12 nm large), stabilized by the enclosure with an amphipathic helical membrane scaffold protein (MSP). A TMP incorporated into nanodiscs is claimed to remain more stable and possibly mono-disperse with the ability to access both sides of the lipid bilayer. These properties may allow performing structural, biophysical and biochemical investigations in a more effective and simpler way. Nanodiscs mimic near-native environments, but their large particle size complicates the spectral analysis of TMPs by solution NMR. Recently this disadvantage was tried to overcome by developing smaller nanodiscs, which are obtained by truncating the MSP. However, applicability to solution NMR needs to be demonstrated for multi-spanning TMPs. In the last chapter, I describe the methodology for incorporating a 3-TM fragment (TM567) of NPY4R into nanodiscs of smaller diameter.

Zusammenfassung

G-Protein-gekoppelte Rezeptoren (GPCRs) zählen zu einer Gruppe von alpha-helikalen transmembran Proteinen (TMPs), die bei unzähligen biochemischen Prozessen der Signalübertragung involviert sind. Diese Rezeptoren kommen in sämtlichen eukaryotischen Organismen vor und weisen in der Natur und vor allem im menschlichen Körper, wo knapp 800 verschiedene Genarten identifiziert wurden, die für TMPs kodieren, eine sehr hohe Diversität auf. Obwohl GPCRs nur rund 1% der Zellproteine ausmachen, sind sie an vielzähligen Signalübertragungsvorgängen beteiligt und somit von grossem pharmakologischen Interesse. Fast 50% aller am Markt erhältlichen Medikamente wirken auf diese Rezeptoren, die als therapeutisches Ziel bei zahlreichen Krankheiten gelten. Die Hauptaufgabe von GPCRs ist es, Signale durch die Zellmembran zu übertragen. GPCRs erhalten dabei einen extrazellulären Reiz von Ionen, Photonen, kleinen organischen Molekülen, Peptiden und sogar ganzen Proteinen (wie zum Beispiel Hormonen). Dies führt zu einer Konformationsänderung der Rezeptoren, welche daraufhin wiederum mit G-Proteinen im Inneren der Zelle in Wechselwirkung treten können. Nach dieser Signalübertragung dissoziieren die G-Proteine vom Rezeptor, spalten sich in zwei Untereinheiten und lösen weiterführende Signalkaskaden aus. Wichtig zu erwähnen wäre hierbei, dass die meisten Medikamente auf die initiale Bindung des Liganden und die einhergehende Konformationsänderung abzielen. Ein besseres Verständnis der Beziehung zwischen Struktur und biologischer Funktion ist daher von höchstem Interesse, um gezielt Medikamente herzustellen, die GPCR-Signalübertragung beeinträchtigen.

Im menschlichen Körper kodieren rund 15% aller Gene für GPCRs, wobei die Funktion der meisten Rezeptoren noch unerforscht ist. Ausserdem sind selbst für GPCRs mit bekannten Funktionen strukturelle Informationen limitiert, da sich ihre Charakterisierung selbst mit modernen Technologien als äusserst anspruchsvoll darstellt. Die Ursache hierfür kann in den Eigenschaften von GPCRs, wie ihre Hydrophobizität, ihr geringes natürliches Vorkommen, ihre hohe Flexibilität und die Schwierigkeit geeignete Membranmimetika zu identifizieren, gefunden werden.

Im ersten Kapitel stelle ich die generellen Eigenschaften dieser Rezeptoren sowie die neuesten Entwicklungen in Bezug auf die Optimierung der Expressionsausbeuten mit heterologen Expressionsstämmen vor. Ausserdem gehe ich auf die Möglichkeiten ein, passende Membranmimetika zu finden, um TMPs zu rekonstruieren und biophysikalisch zu analysieren.

In Kapitel 2 und 3 stelle ich einen 'Divide-and-Conquer' Ansatz vor, um Strukturen G-Protein-gekoppelte Rezeptoren in Lösung zu bestimmen. Dabei wird der menschliche Y4 Rezeptor in 2-3 transmembran grosse Fragmente aufgeteilt, die dann individuell mittels NMR Spektroskopie untersucht werden können. Hierbei untersuche ich systematisch verschiedene Aspekte der Expression der Fragmente in *E. coli* sowie diverse Proteinaufreinigungsstrategien. Im Besonderen vergleiche ich hier die Produktion der Fragmente in Einschlusskörpern mit der unlöslichen ΔTrp leader Sequenz als Fusionsprotein oder mit Mistic, welches das Expressionsprodukt direkt in die Membran dirigiert. Ausserdem stelle ich Methoden vor, die es ermöglichen die Fusionsproteine enzymatisch von den Fragmenten abzuspalten. Die direkte Expression von 2- bzw. 3-TM Fragmenten in Einschlusskörpern diskutiere ich im dritten Kapitel.

Die biophysikalische Analyse von Membranproteinen mittels Lösungs-NMR Spektroskopie ist in letzter Zeit sehr stark fortgeschritten, und im vierten Kapitel beschreibe ich Expression, Aufreinigung und anschließende NMR spektroskopische Charakterisierung aller C-terminalen NPY4R Fragmente. Ausserdem diskutiere ich die Möglichkeit, mit Hilfe von gemeinsamen NOESY Korrelationen, chemische Verschiebungen von kleineren auf grössere Fragmente zu übertragen. Am Ende des Kapitels behandle ich mögliche Erklärungen für fehlende oder nicht zugeordnete Peaks.

In Kapitel 5 beschreibe ich ein spezielles von Sligar und Kollegen entwickeltes Membranmimetikum genauer, das natürlichen Membranumgebungen sehr ähnlich ist. Diese sogenannten Nanodiscs sind diskoidale Phospholipid-Doppelschichten mit einer Grösse von ungefähr 10-12 nm, die durch ein amphipathisches helikales Gürtelprotein (membrane scaffold protein - MSP) stabilisiert werden. In Nanodiscs eingebaute TMPs

sind somit wahrscheinlich stabiler, vermutlich monodispers und von beiden Seiten der Lipid-Doppelschicht zugänglich. Diese Eigenschaften könnten es erlauben, biochemische und biophysikalische Untersuchungen einfacher und effektiver durchzuführen. Obwohl Nanodiscs die natürliche Membranumgebung sehr gut imitieren, bedeutet deren erhebliche Partikelgrösse ein nicht triviales Problem bei der Lösungs-NMR spektroskopischen Untersuchungen der TMPs. In letzter Zeit wurde daher versucht, durch die Verkürzung von MSP kleinere Nanodiscs herzustellen. Die Anwendbarkeit dieser kleineren Nanodiscs für NMR spektroskopische Untersuchungen von mehreren Transmembranhelices enthaltenden TMPs muss jedoch noch demonstriert werden. Im letzten Kapitel beschreibe ich daher die Methodik zur Insertion eines 3-TM grossen Fragments (TM567) von NPY4R in Nanodiscs mit kleinerem Durchmesser.

Chapter I: Heterologous Expression Systems for Eukaryotic Transmembrane Proteins (TMPs) and their Reconstitution into Membrane Mimetics

1. Introduction

Transmembrane Proteins (TMPs) are considered to perform a diverse range of crucial biological functions and thus are of significant pharmacological importance. TMPs comprise only 1% of the total cellular proteins but pharmacologically TMPs are the targets for approximately 50% of drugs in the market and in research[1, 2]. Structural characteristics of alpha-helical (as in contrast to beta-barrel TMPs) TMPs resemble a single transmembrane helix or a bundle of transmembrane helices existing independently or interacting with other partners within the host cell membrane environment. TMPs act as intermediary signaling networks by interacting with broad range of extracellular ligands and subsequently undergoing conformational changes that generate wide variety of stimuli directing intracellular responses. Thus from a human curiosity point of view, it is obvious to conquer their structural implications for understanding biological mechanism[3, 4]. In both Prokaryotes and Eukaryotes, approximately 25% of all genes code for TMPs and especially in humans 15% of these TMPs is G-protein coupled receptors (GPCRs)[2, 5]. Out of the vast majority of TMPs identified, function for most of them has not yet been assigned and for those the function has been assigned they lack high-resolution structural information. It shouldn't be construed that the minimum availability of high-resolution structures of TMPs meant lack of efforts by the scientific community but it's because of the tremendous challenges displayed by their hydrophobic nature, low natural abundance, high flexibility and the lack of an ideal membrane setup.

Again from all the issues mentioned above for studying TMPs in detail, especially 2 of them are very critical: Obtaining sufficient quantities and selecting an ideal membrane mimetic. Thus in this chapter, I shall address the progress in troubleshooting developments or the alternatives available in overcoming the above 2 critical challenges.

2. Recombinant expression of proteins: After the development of recombinant technology in the 1960's, *in vitro* studies on all types of proteins have drastically increased and there is a surge on the number of crystal structures being deposited into the

PDB. But the number of entries for TMPs compared to soluble proteins is insignificant because of the unavailability of a perfect expression system that can deliver sufficient quantities. However, during the course of exploring new expression systems or new methodologies/procedures for expressing TMPs, many host systems have come up with distinctive characteristics solving the problem of 'protein yields' to a promising extent (Fig 1).

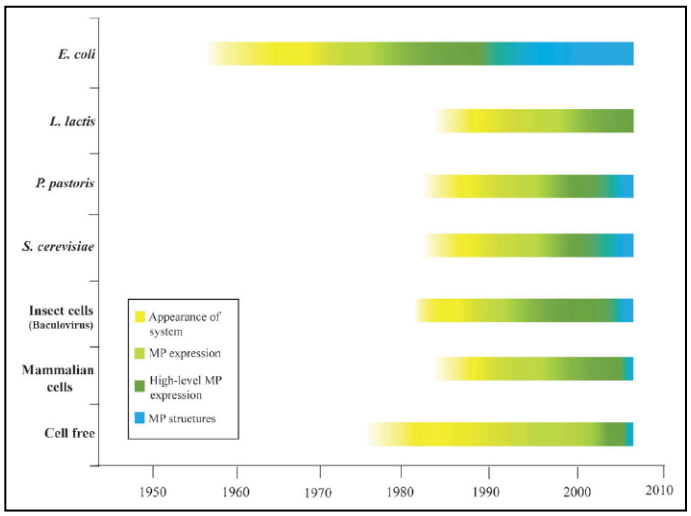


Figure 1. Development progress of different TMP expression systems. Well-known TMP expression systems are compared with regards to their emerging applications for protein expression (yellow), initial applications to TMP production (light green), improvement for preparative-scale TMP production (dark green) and first appearance of TMP structures (blue)[6].

These host systems should be amenable for large-scale production resulting in at least milligram quantities of purified, homogenous TMPs for high-resolution structure determination (X-ray, NMR or EM). A heterologous system chosen to express TMPs again has certain limitations[7] and some of them are: Toxicity to the host, low-levels of expression, short half-lives before degrading[8], failure to produce properly folded states and lack of certain post-translational modifications (Fig 2).

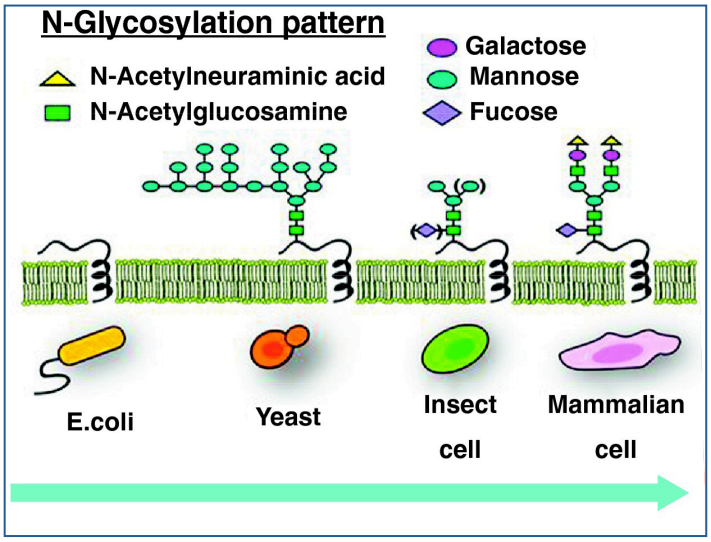


Fig 2. Glycosylation of TMPs when expressed in different hosts[9]. In *E.coli*, N-Glycosylation of TMPs is completely absent, while in mammalian cells TMPs are completely glycosylated.

Recombinant protein overexpression requires 3 crucial elements: Suitable vector, expression host and ideal expression conditions (adding inducer at the correct cell density, inducer concentration and post-induction temperature). Mix and match of the suitable above combinations maximize the quantity and quality of the protein. In any experimental setup for the overexpression of TMPs in a heterologous host, it's difficult to predict a standard combination because of the diverse nature of TMPs having distinct physico-chemical properties. Even after overexpression of TMPs, it's difficult to predict if they can be purified or if they are in the active state or even if they are suitable for high-resolution (X-ray, NMR) structure determination. Thus for a better end result, it's always good to test diverse heterologous systems. In this chapter the focus is mainly on 4 aspects:

1. Recent techniques developed for high-throughput cloning strategies
2. *E.coli* as expression system
3. Heterologous expression systems other than *E.coli* for TMP overexpression and purification
4. Reconstitution of purified TMPs in an ideal membrane mimetic.

3. Cloning Strategies:

The first step when overexpressing TMPs in a heterologous system is cloning. Traditional cloning of the gene of interest (GOI) using restriction enzymes and ligase into multiple expression vectors is time consuming. Thus, alternative cloning strategies have come up in the last few years that allow for quick cloning of hundreds of genes generating multiple constructs simultaneously[10]. Some of these cloning technologies are: Gateway cloning, Creator and the fragment exchange (FX) cloning[11-16]. These cloning strategies also take into account addition of various fusion tags that could be effective in overexpression, folding and purification. Fusion tags can be placed at either the N- or the C-terminus of a target protein depending on the purpose it serves and most often it is difficult to know in advance, but can be determined experimentally both the position and the type of fusion.

Once the design of the expression vector is ready, overexpression of the TMPs can be tested in suitable prokaryotic and eukaryotic systems, keeping in mind the pros and cons of each of them towards functional protein expression[12, 17-22].

Prokaryotic expression systems are:

a) *Escherichia Coli*
b) *Lactococcus lactis*
c) *Rhodobacter sphaeroides*

Eukaryotic expression systems are:

d) Yeast
e) Insect cells
f) Mammalian cells

4. Prokaryotic expression systems

4.1. *Escherichia Coli*

To date, *E.coli* is the most extensively used expression host for producing recombinant proteins. It's ideal characteristics such as ease of use, low cost and short generation times resulted in expressing almost all types of recombinant soluble and TMPs. However,

expression of TMPs in *E.coli* presents many disadvantages: Many TMPs such as GPCRs often are not expressed to sufficient quantities: In most cases, expressed TMPs are not properly folded or result in formation of insoluble aggregates such as inclusion bodies, or they lack post-translational modifications. Since the *E.coli* system has been well-characterized, recent developments showed significant improvement in TMP expression[23, 24]. Strains such as C43, C41 & Lemo 21 emerged as tolerant strains for toxic TMPs in *E.coli*[7, 25]. Also, TMP production can be enhanced by incorporation of various fusion tags such as GFP[26], MBP, GST, NusA[6], ΔTrp or Mistic[27], etc. and their influence on TMP expression is as follows:

Fusion tag	**Size**	**Function in *E.coli* cells**
$His_{(6\text{-}10)}$ - tag	1 kDa	Most common affinity purification tag
Maltose binding protein (MBP)	43 kDa	Enhances solubility and folding
N-utilization substance (NusA)	54.8 kDa	Enhances solubility and folding
Thioredoxin (Trx)	12 kDa	Enhances solubility and avoids inclusion body formation
IgG domain B1 of *streptococcus* Protein G (GB1)	6 kDa	Enhances solubility and usually used in protein preparations for NMR
Trp leader sequence (ΔTrp)	14 kDa	Diverts TMPs into inclusion bodies
GST	26 kDa	Enhances solubility and acts as a purification tag
Small ubiquitin modifier (SUMO)	~11 kDa	Enhances solubility and stability
Mistic	13 kDa	Enhances overexpression and solubility of TMPs.
Ketosteroidoisomerase (KSI)	~15 kDa	Diverts TMPs into inclusion bodies

A special emphasis shall be put on Mistic, a 13 kDa protein originally from *Bacillus subtilis.* Mistic*,* when produced recombinantly in *E.coli*, was postulated to directly associate with the inner membrane without the need for recognition by the Sec translocon machinery. Mistic's unusual nature to spontaneously associate with the membrane

facilitated successful insertion of various TMPs into *E.coli* membranes when fused to Mistic as an N-terminal fusion tag[27-31].

Tags might not be suitable or even necessary for TMP expression in certain strains. During the reconstitution process, some of the fusion tags can interfere with the active site that can result in altered structure and function of the target protein. Thus cleavage recognition sites (enzymatic or chemical) can be engineered into the expression construct to cleave the fusion tags at different stages of purification. Few commonly used recognition sites are:

Protease/chemical recognition site	**Recognition Sequence**
Enterokinase	DDDDL|
Factor Xa	I D/E GR|
TEV protease	ENLYFQ|G
C3 protease	LEVLFQ|GP
Thrombin	LVPR|GS
CNBr	X M|X
Signal peptidase	AXA|

4.2 *Lactococcus lactis*

L. lactis is a non-invasive, non-pathogenic, Gram-positive bacterium, which is similar to food-grade lactic acid fermenting bacteria. Because of the wide-ranging developments in genetic engineering tools, *L. lactis* is also extensively used for the large-scale production of recombinant TMPs[32]. TMP expression in *L. lactis* is done with the Nisin-Inducible Controlled gene Expression (NICE) system where Nisin, an antimicrobial peptide, acts as an inducer for the genes positioned under the control of promoter $P_{nisA.}$ Several eukaryotic TMPs have been expressed using the NICE system in *L. lactis*[33-35]. The main advantage of using *L. lactis* over *E. coli* is that the TMPs expressed were always directed into membranes and not into inclusion bodies[36, 37]. The other advantage is the presence of

a single membrane that can help in performing activity studies by directing ligands or inhibitors on the whole cells.

4.3 *Rhodobacter sphaeroides*

R. sphaeroides belongs to the class of purple, non-sulphur photosynthetic bacteria. The main advantage of this expression system is the presence of an increased membrane surface area, which is a major limitation for most of the hosts. It consists of chromatophores in the photosynthetic apparatus made up of the pigment-protein complex (light-harvesting complex and reaction centers), situated in the projections of the cytoplasmic membrane. These bacteria synthesize large amounts of photosystems on exposure to light and/or reduced oxygen availability resulting in the production of more chromatophores that leads to the increase in membrane surface availability[38]. This property of increase in intracytoplasmic membrane surface can be exploited for the production of TMPs, which was already being implemented[39].

5. Eukaryotic expression systems

5.1 Yeast

In general, yeast has numerous advantages as host system for the production of eukaryotic TMPs. Similar to bacteria, yeast systems are well characterized, can be inexpensively cultured in large volumes, their plasmids can be straightforwardly genetically engineered but unlike bacteria, yeast systems have an additional advantage of possessing protein processing machinery for adding post-translational modifications (PTMs) to the expressed proteins that are similarly observed in mammalian cells. Many yeast systems have been tested and engineered for eukaryotic protein production but *Saccharomyces cerevisiae* and *Pichia pastoris* are the two main yeast systems, which are widely used for TMP production. Yeast system *S. cerevisiae* is the most extensively characterized strain for TMP production and in general has vast variety of sub-strains to work with, whereas the yeast system *P. pastoris* mediates a stringently coordinated inducible expression system[18]. From an NMR point of view it is worth to mention that perdeuterated proteins can be produced in *Pichia Pastoris*.

5.1.1 ***P. pastoris*****:** It has the capability to utilize methanol as an exclusive carbon source. Methanol metabolism occurs in peroxisomes and the first reaction is the conversion of methanol to formaldehyde and hydrogen peroxide by alcohol oxidase[40]. Thus in the presence of methanol, peroxisomes significantly enlarge in size due to high expression levels of enzymes that are necessary for complete methanol metabolism (Fig 3). *P. pastoris* mainly expresses 2 alcohol oxidase genes (AOX1 & AOX2) and they assemble as ~600 kDa homo-octamers. While 10% of total oxidase activity is from AOX2, the majority is from AOX1. Thus expression levels of AOX1 increase from trace amounts to ~30% of total intracellular protein in presence of methanol. Thus the methanol-inducible *AOX1* promoter is tightly regulated and is an ideal system for heterologous protein expression. Exploiting the above concept, *P. pastoris* expression vectors were designed featuring the AOX1 promoter for producing many heterologous soluble and TMPs including GPCRs. In one such high-throughput effort, 100 different types of mammalian GPCRs were expressed in *P. pastoris,* of which 52 were investigated for ligand binding and 20 were screened for high expression yields[41].

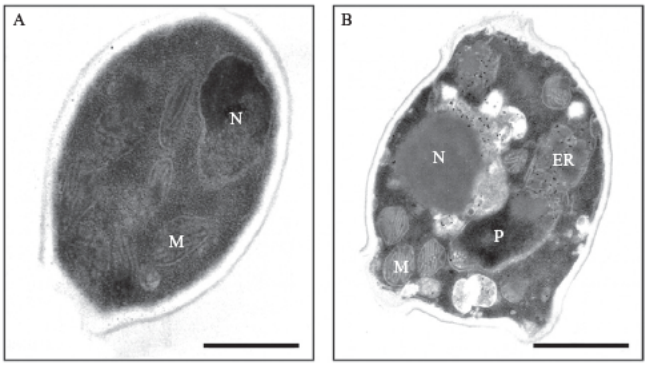


Figure 3. Electron micrographs showing morphological transformations in *P. pastoris* before (A) and after (B) methanol induced β2 adrenergic receptor protein expression[41]. (A) Medium without methanol. (B) Same clone as in (A), but induced with methanol. N - nucleus; M - mitochondria; P - prominent peroxisome following elevated H_2O_2 during methanol metabolism; ER - endoplasmic reticulum with expressed receptor.

5.1.2 ***S. cerevisiae*****:** This is the first yeast system used for the expression of recombinant proteins. The most frequently used promoters are GAL1 and GAL10, which are galactose inducible. It is observed that 3 crucial parameters should be followed for the production

TMPs such as GPCRs to obtain permissible yields: (i) use of a robust promoter such as GAL1 and co-expression of its transcriptional transactivator; (ii) tagging the N-terminal end with a yeast signal sequence (N-terminus of STE2 receptor) that favors insertion into membranes (iii) induction should be carried out in presence of an antagonist that could act as a structural stabilizer. However, subsequent studies showed that the above methodology doesn't work for all GPCR expression. The most successful observation was for human β_2-adrenergic receptor with yields of 115 pmol of receptor per milligram of total membrane protein[42]. On the contrary, only 20 fmol/mg of membrane protein was observed for the production of μ1-muscarinic receptor[43]. Moreover, following the above described conditions for human β_2-adrenergic receptor, but swapping the cultures from lab-scale to fermenters (preparative scale) decreased the yields to 36 pmol/mg of membrane protein[44].

Recombinant TMP expression in yeast systems drives the focus on mainly towards PTMs. PTMs mainly comprise of the glycosylation patterns and clearly there is huge difference compared to those in mammalian cells. Hyperglycosylation is generally observed in *S. cerevisiae* whereas in *P. pastoris* PTMs are at optimum levels. Thus the level of glycosylation for expressed proteins in yeast can not only effect activity but also trafficking, folding, stability and degradation[45]. Recently, yeast systems have been engineered so that they can glycosylate recombinant proteins with the patterns observed in mammalian systems and these modified yeast strains could in the future become the potential strains of choice for all eukaryotic recombinant protein expression[46]. Another major difference to be considered when expressing TMPs in yeasts is the biochemical nature of the membrane environment. Both yeast systems have significant variations from the mammalian cells in the lipid composition that make up the membranes[11, 47]. Endogenous cholesterol found in the mammalian cell membranes is totally absent from yeast membranes, which might be crucial for the functionality of the expressed protein as is in the case of SERT1 serotonin receptor [48].

5.2 Insect Cells

In general, insect cells are easy and inexpensive systems to handle compared to mammalian systems like CHO, HEK 293 or COS cells, particularly for scale-up, but are

expensive to culture compared to yeast and bacteria. Principally, the system relies on baculovirus (recombinant) infection of insect cell lines (Sf9, Sf21 or High Five). Thus this is also called baculovirus expression system (BVES). After the viruses infect the insect cells (host), they multiply along with the expression of recombinant protein until they reach a certain viral titer. In the last decade, many improvements have been executed to generate recombinant baculoviruses and one of them comprises the Bac-to-Bac system (Invitrogen), which utilizes site-specific transposition in *E. coli* rather than homologous recombination in insect cells (fig 5)[49, 50]. The polyhedrin or p10 late promoter normally drives the gene expression (fig 4)[51, 52]. Recently, a similar system (BacMam, Invitrogen) has been established to allow baculovirus-based expression in mammalian cells.

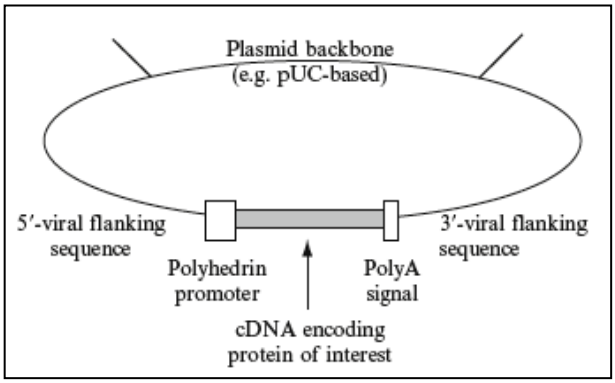


Figure 4. Representation of a simple baculovirus transfer plasmid[125].

BVES is extensively used for eukaryotic protein expression in insect cells as a compromise between mammalian expression systems (stably or transiently transfected) and the bacterial expression systems. Although, BVES is more costly and laborious than expression in *E. coli*, it is better suited for eukaryotic proteins (especially TMPs) because of (i) similar codon usage rules, (ii) better expression levels (iii), better folding and fewer truncated proteins and (iv) allowance for post-translational modifications that are close to mammalian systems. A notable fact is that, some of the post-translational modifications generated are not identical to those found in mammals (for example glycosylations), but they are closer than those produced by bacteria, or even yeast[49].

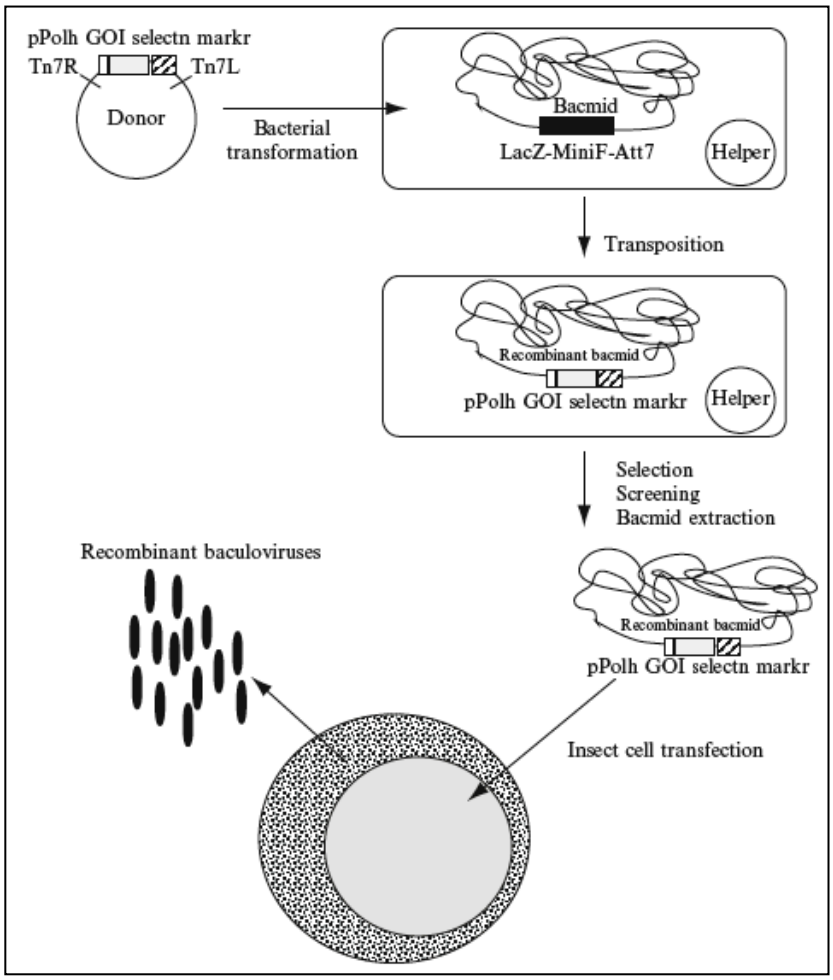


Figure 5. Generation of recombinant baculoviruses from vectors obtained through transposition between a transfer plasmid and a bacmid[53].

5.3 Mammalian Cells

Mammalian cells, and in particular those of human origin, undoubtedly offer the most native biochemical composition for the expression of TMPs that are linked to human physiology and disease. Until the last couple of years, their use in structural biology had been limited because of the practical difficulties in culturing and transfecting large quantities of cells and more crucially the costs endured in maintaining them. Conventionally, TMP expression in mammalian system depended on obtaining stable transformants or on the use of viral vectors. Most commonly used viral vector for TMP expression is the Semliki Forest Virus (SFV) and in several cases, promising quantities of homogeneous, functional protein has been obtained[17, 54, 55]. However both, the generation of stable transformants and the usage of viral stocks require expensive culture media and

thus are technically challenging. But, in the last decade, acceptably successful and relatively cheaper large-scale transient transfection (LSTT) techniques have evolved with the development of serum-free media as an alternative expression method[56-58]. In recent years, efforts were put into increasing transfection efficiency for higher expression yields by optimizing various media components and culture conditions[57, 58].

The Structural Proteomics IN Europe (SPINE) group has carried out detailed investigation on expression characteristics of all kinds of proteins using LSTT and concluded that the highest overall expression was observed for secreted proteins relative to other types of proteins[59]. SPINE group also investigated protein expression using baculoviruses, and concluded that this method gave better yields for cytosolic and TMPs. Other investigations also revealed that the mammalian cells are best-suited systems for functional TMP production[60].

A remarkable approach comprises the use of baculoviruses as mammalian expression vectors[61, 62]. The concept of baculoviruses infecting mammalian cells was known in the mid 1990's, but scientists were not able to replicate the same phenomenon with other viral vectors for avoiding safety concerns. Recombinant shuttle baculoviruses are generated using the shuttle vector pfastbacmam-1, in which the traditional polyhedrin promoter is substituted with a cytomegalovirus (CMV) promoter[63], and this system has since been applied to a wide variety of cell lines, mainly for receptor binding assays[64-66] but more recently also for large-scale protein production [67]. Lately, a single shuttle vector comprising both the CMV and pol10 promoters are under widespread usage, allowing expression in both mammalian and insect cells[50]. One of the drawbacks posed with the usage of induced promoters in transient transfection is that gene expression is constitutive, and that protein overexpression could easily cause cell death due to toxicity. To overcome this, a vital development has taken place in which protein overexpression can be tuned to levels that avoid toxicity[68-71].

6. Comparison of different expression systems

The choice of selecting an expression system is highly empirical with regards to functional TMP production. Considering a few realistic examples:

- Functional SERT1 a serotonin transporter is expressed at low levels in yeast and insect cells compared to mammalian cells [48].
- Functional AQP4 suitable for crystallization was produced in high quantities in insect cells compared to yeast and CHO cells[72].
- Recombinant rhodopsin has been produced in almost all expression systems, but in comparison insect cells had the highest levels of functional expression while yeast had the lowest expression levels[73-76]. Moreover, only mammalian COS cells produced the glycosylated rhodopsin, which is analogous to the bovine rhodopsin and was successful in crystallization[77].

Generally, it's the target protein, which determines whether a particular host is ideal or not for expression, and thus extensive screening is required to determine the best host system and its best strain under ideal conditions[17]. Due to the wide range of recent developments in all the expression systems, several potential comparative platforms have emerged especially for mammalian and insect-cell systems that are conventionally tedious and time-consuming[11, 78].

Table 2. Comparison of major expression systems for the production of functional TMPs based on the type of recombinant technique followed including the advantages and disadvantages[32].

Expression system		**Advantages**	**Disadvantages**
E. coli	Plasmid	Low-cost	Problems in folding
		Easy to culture	Low yields
		Quick construct generation	No PTMs
		Compatible with most media for labeling	
Yeast	Plasmid	Inexpensive and can be grown to a high density	Non-native PTMs/hyperglycosylation (*S.*

			cerevisiae)
		Easy to culture	Non-native membrane environment
		Quick construct generation	Poor expression of some proteins
		Eukaryotic processing	Poor folding
Insect cells	General	More native membrane environment than yeast	Costly
		Somewhat robust culture requirements	Non-native PTMs
		Good track record in producing functional proteins	Non-native membrane environment
	Baculovirus	Well-established protocols	Time needed to generate virus
		Easier titration of viruses	Cell lysis, safety concerns
	Transient transfection	Quick construct generation	Costly
			Amount of DNA required
	Bacmam	Cross-over	Necessity for two different types of cells
			Lysis in insect cells
Mamma -lian cells	General	Native membrane environment	Costly
		Native secretory/post-translational pathways	Technical requirements
		Good track record in producing functional proteins	
	Stable integration	Consistent expression levels	Time needed to establish
		No need to generate a vector	Instability in integration
	Semliki Forest Virus (SFV)	Successful in expression-screening studies	Necessity for helper RNA
		Efficient infection	Technically demanding
	Transient transfection	Rapid	Amount of DNA required
			Price of reagents at large scale
	Bacmam	Lack of degradation in mammalian cells	Necessity for two different types of cells
		Cross-over	

7. Cell-free expression system (CFES)

CFES for analysis of various proteins has been used for several decades[79, 80]. The main drawback of this system during its early days was the low yield of target proteins. Reticulocyte lysate or wheat germ extract was the first mode of medium used for CF protein synthesis, especially for analyzing specific antibodies that are radiolabelled by autoradiography or by western blotting[80]. In the late 1990s, significant improvements were made for obtaining higher expression yields in the medium composition of both an *E. coli* and wheat germ derived cell-free systems. In the last decade, CFES for especially TMP production has undergone significant progress due to the advances made in the CF ingredients by several marketing companies (e.g. Novagen, Roche, Qiagen, Invitrogen)[81-84]. Thus several commercial cell-free extracts are available in the market based on *E. coli* strains and wheat germ lysates for preparative production of proteins. Some of the advantages of CFES over cell-based systems are:

- Expression of cell-toxic proteins
- Production of difficult-to-express proteins on preparative scales quickly and economically.
- Simultaneous expression of multiple proteins in one reaction
- Easier incorporation of non-natural aminoacids
- Easier isotopic labeling, in particular amino-acid specific labeling with less scrambling
- Feasibility to use suitable membrane mimetics for TMP production
- Other additives that could be added are metal ions, cofactors, binding partners/ligands, etc.
- In some cases, PCR amplified product with the gene of interest is sufficient to act as template for protein synthesis. Thus doesn't require cloning into vectors.

Cell-free Reaction: It works in two different ways: First is the **batch mode**[82] in which the translation reaction is allowed to undergo in a vessel that has all the necessary ingredients. This type of reaction can only last for 2-3 hrs as the initially supplied energetic sources, nucleotides, metal ions and cofactors start to get either depleted or

degraded. This mode of reaction can produce up to 500 µg/ml protein and sometimes up to 1 mg/ml protein in exceptional cases with proper optimization.

The second method is the **dialysis mode** cell-free reaction[85]. Here, the translation reaction is performed in a reaction compartment that is separated by a dialysis membrane from the reservoir chamber (10-20x larger than reaction compartment) consisting of lower molecular weight ingredients. In this setup, reactions usually last for 20-24 hrs resulting in higher protein yields. Protein yields up to 5 mg/ml were reported with the *E. coli* system for both the marketed (Roche, Invitrogen) and as well as with the self-prepared ingredients according to published protocols. With the wheat-germ system, protein yields reported were less, usually 1 mg/ml but this system provides stable reactions that can run for several days[86, 87].

Basic ingredients required for a cell-free reaction are *E. coli* strain's cytosolic extract (BL21DE3, BL21 CodonPlus, A19, etc.), buffers such as Tris-acetate (pH 8.2), HEPES-KOH (pH 7.5) and other chemical compositions such as KOAc, $Mg(OAc)_2$, K-glutamate, NH_4OAc, DTT, NaN_3, PEG8000, β-ME, NTPs (ATP, GTP, CTP, UTP), 20 amino acids, 3',5'-cyclic AMP (cAMP), Folinic acid, creatine phosphate, phosphoenol pyruvate (PEP), creatine kinase, pyruvate kinase, T7 RNA polymerase and total tRNA mix from *E. coli*. In addition to the above components, TMP production in active form requires membrane mimetics such as detergents, lipids and nanodiscs[84, 88-90]. Extensive optimization is required for the translation reaction to take place as excess membrane mimetics could affect the activity of enzymes in the reaction mix. Recent studies have shown efficient production of functional TMPs including GPCRs[91-95].

8. Membrane Mimetics

Although obtaining sufficient quantities of TMPs has progressed drastically by using tailored heterologous systems, their incorporation into membrane mimetics in functional form is still a major bottleneck. Lipids in cell membranes not only act in embedding TMPs but also regulate their function by their dynamic rearrangements[96]. TMPs sometimes are embedded in unique lipidic environments native to certain cells[97-99] and thus *in vitro* experiments carried out should satisfy the following properties to support comprehensive biological characterization:

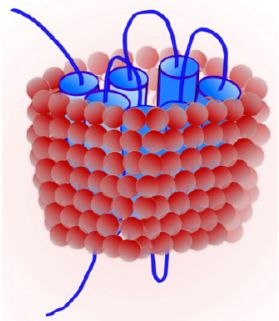
- The selected membrane environment should support ideal folding of TMPs
- Membrane environment surrounding TMPs should provide appropriate access to its binding partners (peptides, cofactors, metal ions, ligands)
- Membrane environment should regulate oligomeric state of TMPs based on functional significance

Although detergent micelles, liposomes (artificial lipid bilayers, bicelles) and nanodiscs are reported to provide native-like environments for TMPs, none of them fulfill all the above-enlisted criteria. For understanding the incorporation of TMPs into membrane mimetics, we will first consider the most common types available for functional reconstitution *in vitro* and then will analyze their advantages and disadvantages[100].

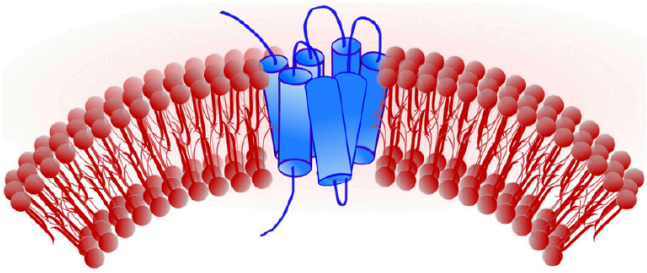
Basically, there are 5 types of membrane mimetics widely used by most labs. These are:

i. Detergent or mixed micelles
ii. Bicelles
iii. Unilamellar lipid vesicles
iv. Nanodiscs
v. Amphipols

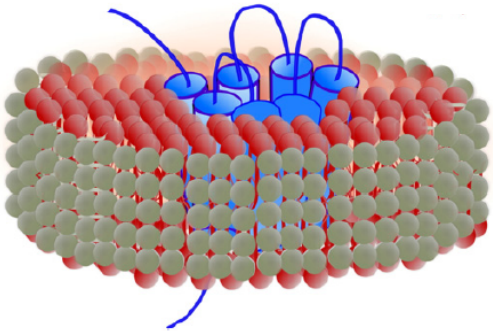
An imaginary view of each membrane mimetic with the TMP integrated is depicted in the following figure (Fig 6).



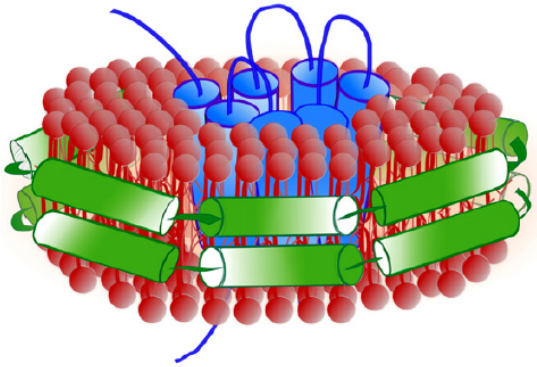
6a. TMP in a detergent micelle



6c. TMP in unilamellar lipid vesicles



6b. TMP in bicelles



6d. TMP in nanodiscs

8.1 Detergent or mixed micelles: These are the most common membrane mimetics used for *in vitro* studies of purified TMPs[101, 102]. Micelles or mixed micelles are formed by adding a single type of detergent, a mixture of detergents or a combination of detergents and lipids in water at certain levels of concentration (Figure 6a). Because the TMPs consist of both hydrophilic and hydrophobic regions, a stable environment is provided to them by the detergents, which are amphipathic molecules. Thus detergents prevent precipitation of TMPs in water. Wide ranges of detergents are available for solubilizing TMPs, but a selected detergent is considered ideal if it is mild enough to prevent instant denaturation of the protein. For example, sodium dodecyl sulfate (SDS) solubilizes TMPs effectively but often in a denatured mode, and thus SDS is not the best choice. On the contrary, *n*-octyl-β-d-glucopyranoside (OG)[103], cholate and dodecyl-β-d-maltoside (DDM)[104] act as mild detergents and are used with greater success. Crystal structure of one of the TMP rhodopsin was first solved in presence of the mixed micelles composed of heptanetriol and nonyl-β-d-glucoside[105, 106].

Micelle formation is a thermodynamically driven process to shield the hydrophobic tail of detergents from getting exposed to water while maintaining the interactions of polar head groups with water. When a detergent is added to water, they disperse as monomers upto certain concentrations and when these concentrations are exceeded, they tend to aggregate forming micelles. The concentration at which the detergent monomers tend to aggregate is called as the Critical Micelle Concentration (CMC). Depending on the type and length of the hydrophobic tail and to a certain extent the type of head group involved, CMC values change. The number of detergent monomers that associate to form the micelle is called as the Aggregation number. This also depends on the type of detergent monomer based on similar properties such as the length of the hydrophobic tail and the head group involved. Most commonly used detergents in structural biology are illustrated in table 3.

Basically, detergents that solubilize TMPs and suitable for NMR fall into two categories: **(a)** Detergents with apolar tails – dodecylphosphocholine (DPC), sodium dodecylsulfate (SDS), N-lauroylsarcosine, lauryldimethylamine oxide (LDAO), decyl- and dodecyl-maltoside (DM, DDM) and octylglucoside (OG). Most commonly used tailed detergents suitable for NMR are dihexanoylphosphatidylcholine (DHPC) with two short tails and lyso-glycerophospholipids (phospholipids without their sn-2 acyl chain). Tailed detergents with uncharged headgroups were not used extensively on TMP studies using NMR because they yield poor-quality spectra. It is speculated that the micelle monodispersity with these detergents is not well preserved at high concentrations suitable for NMR studies, resulting in longer correlation times for the TMP- detergent complexes. Tailed detergents with charged head groups were extensively used in studying TMPs using NMR because they yield high-quality spectra. Some of these detergents such as DPC, SDS, LDAO and N-laurylsarcosine should be used carefully as they have the tendency to destabilize folded TMPs. Widely used detergents in this category are the those with short-chained phospholipids (DHPC)[126] and lysophospholipids[127] (LMPC, LMPG, LPPG – table 3). These are zwitterionic in charge, resulting in high water solubility and the micelles mostly exist in monodisperse form. A minor disadvantage is that these detergents are very mild in nature to solubilize TMPs. **(b)** These are rigid amphipathic molecules similar to bile salt-derived detergents or amphipathic peptides.

Most common example is CHAPS[128]. In addition to solubilizing TMPs, CHAPS is also capable of stabilizing small exposed hydrophobic patches present in few water-soluble proteins.

Detergent	Charge	Molecular Weight	Critical micelle concentration (mM)	Aggregation number
CHAPS	Zwitterionic	615	6–10	4–14
CHAPSO	Zwitterionic	631	8	11
β-Decylmaltoside	Nonionic	483	2	104
β-Dodecylmaltoside	Nonionic	511	0.2	110–140
β-Octylglucoside	Nonionic	292	19–25	90
Dihexanoylphosphatidyl choline (DHPC)	Zwitterionic	454	15	19–35
Diheptanoylphosphatidyl choline	Zwitterionic	482	1	42–200
Decylphosphocholine	Zwitterionic	323	11	ND
Dodecylphosphocholine (DPC)	Zwitterionic	352	1.5	50–60
Tetradecylphosphocholine	Zwitterionic	380	0.12	ND
Lyso-lauroylphosphatidyl choline	Zwitterionic	440	0.5–0.7	70[a]
Lyso-myristoyl phosphatidylcholine (LMPC)	Zwitterionic	468	0.04–09	100[a]
Lyso-palmitoyl phosphatidylcholine (LPPC)	Zwitterionic	496	0.004–0.008	140[a] 186
N-Dodecyl-*N*,*N*-dimethylglycine (empigen BB)	Zwitterionic	272	2	ND
Lyso-myristoyl phosphatidylglycerol (LMPG)	Anionic	478	0.2–3 (pH/salt-dependent)	ND
Lyso-palmitoyl phosphatidylglycerol (LPPG)	Anionic	506	0.02–0.6 (pH/salt-dependent)	125
N-Lauroyldimethyl amineoxide (LDAO)	Zwitterionic	229	2	69–73
N-Lauroylsarcosine	Anionic	293	*ca* 15	ND
Sodium dodecylsulfate (SDS) (precipitates below 15 °C)	Anionic	288	1–7	62–101
Lithium dodecylsulfate (soluble below 15 °C)	Anionic	272	Similar to SDS	Similar to SDS
Digitonin	Anionic	1229	<0.30	60

Table 3. Most widely used detergents for structural biology purpose and their properties[129]. ND – not determined.

8.2 Bicelles: Upon removal of detergent from an incorporated TMP present in a mixture of lipid/detergent micellar environment made-up of both short- and long-chain lipids, a bicellar environment is formed (figure 6b)[130]. Bicelles are nothing but fragments of lipid

bilayer with the outside border stabilized by short-chain lipid or detergent molecules [107-109]. It is interesting that the size of the bicelle can be adjusted depending on the ratios of short- to long-chain constituents. The resulting sizes could range from 10-30 nm and sometimes up to 80 nm is possible upon adjustments for reconstituting multiple GPCRs. Important point to be noted here is, specific lipid ratios should always be maintained for the bicelle entity to persist. Khorana and his coworkers were successful in using the bicelle system to study kinetics of opsin-retinal interactions that results in forming the GPCR rhodopsin[110]. Opella and his coworkers successfully incorporated the CXCR1 chemokine receptor into bicelles, and obtained encouraging NMR spectroscopic results[108, 111]. Recently, the Y2 receptor (GPCR) has been successfully reconstituted into bicelles for solid-state NMR studies[112].

8.3 Unilamellar lipid vesicles: It was observed for TMPs such as GPCRs that they could be reconstituted into unilamellar lipid vesicles from the detergent-solubilized form. These vesicles are generally prepared by solubilizing TMPs and phospholipids in presence of cholate (detergent), followed by detergent removal with the help of bio-beads or by SEC and eventually collecting the separated vesicles (figure 6c). The best-studied TMP that has been successfully reconstituted in unilamellar lipid vesicles is the β_2-adrenergic receptor along with its G_s-subunit[113]. These vesicles are considered to be more advantageous than micelles because of their resemblance towards native-like environment. Recently, Grisshammer and his coworkers reconstituted neurotensin receptor into lipid vesicles and successfully carried out solid-state NMR studies[114].

8.4 Nanodiscs: These are also denoted as high-density lipoprotein (HDL) particles or as nanoscale apolipoprotein bound bilayers (NABB) and present a perfect example of how creative scientists considered using HDLs to incorporate TMPs. Basically a nanodisc is a precisely structured bicelle entity stabilized by 2 molecules of membrane scaffold protein (MSP). The bicellar/bilayer entity in a nanodisc is a lipid bilayer that can incorporate TMPs in its center. MSPs are helical repeat proteins with a hydrophilic and a hydrophobic surface, which can wrap around the lipid bilayer's hydrophobic surface entity resulting in a sturdy bicellar unit called the nanodisc (figure 6d). Because of the nature of MSP function, it is also referred to as a belt protein.

In 2009, Sligar and his coworkers were successful in designing nanodiscs of defined diameter by manipulating the number of helical repeats in the MSP. This resulted in discs with varied sizes ranging from 9-13 nm[115, 116]. Reconstitution of TMPs into nanodiscs offers numerous advantages:

- Major advantage is the absence of a detergent that is responsible for generating a dynamic system.
- Since the lipid bilayer within the nanodisc is not dynamic, free or excess lipid molecules in the solution do not distort the folded TMP in the compact nanodisc.
- In most cases, nanodiscs with TMPs incorporated are considered to be stable for weeks to months whereas it is only days for micelles or bicelles.
- Because of the absence of an enclosed topology, intracellular and extracellular sides of the receptor are easily accessible and amenable for ligand binding studies.
- Because TMPs within nanodiscs have limited conformational flexibility, they act as promising vehicles for structure determination by NMR.
- These entities can be implemented into cell-free expression systems for the direct co-translational reconstitution of TMPs[131].

Some of the successful illustrations with the usage of nanodiscs are:

- Sligar and his coworkers were able to incorporate purified β_2-adrenergic receptor into nanodiscs[117].
- Kobilka and his coworkers revealed that the above incorporation could activate G-proteins[118, 119].
- Rhodopsin incorporated into nanodiscs showed light-induced activation of the receptor and also activation of the G-protein transducin or arrestin[120-123].
- Recently, Sakmar and his coworkers successfully incorporated purified CCR5 chemokine receptor into nanodiscs[124].

The size of nanodisc entities with the presence of reconstituted TMP is estimated to be around 150-200 kDa. These large sizes have in the past, discouraged NMR structural biologists, but recently this system has regained hope after showing promising results with encouraging NMR spectra for medium-sized TMPs[132-133]. Many TMPs using

nanodiscs are at the moment under investigation by NMR, which shows the likely interest of structural biologists to study TMPs in genuine membrane mimetics.

8.5 Amphipols: These are one of the novel classes of model membranes used for studying TMPs by NMR. Chemically, these are amphipathic polymers that can act as alternatives to detergents by keeping the TMPs soluble in water under mild conditions. These are first used as a medium to study OmpA, which gave high-quality NMR spectra[134]. In this example, amphipol A8-35 was added to the detergent solubilized OmpA, following which the detergent was removed using polystyrene bio-beads. OmpA remained in solution trapped by A8-35, which on NMR analysis gave convincing spectra for backbone determination. This study proved that amphipols could replace detergents to study TMPs using solution-state NMR.

9. References

1. Bernaudat, F. et al. Heterologous expression of membrane proteins: choosing the appropriate host. *PLoS One* **6**, e29191 (2011).
2. Von Heijne, G. The membrane protein universe: what's out there and why bother? *J Intern Med* **261**, 543-557 (2007).
3. Lacapere, J. J., Pebay-Peyroula, E., Neumann, J. M. & Etchebest, C. Determining membrane protein structures: still a challenge! *Trends Biochem Sci* **32**, 259-270 (2007).
4. McLuskey, K., Roszak, A. W., Zhu, Y. & Isaacs, N. W. Crystal structures of all-alpha type membrane proteins. *Eur Biophys J* **39**, 723-755 (2010).
5. Wagner, S., Bader, M. L., Drew, D. & de Gier, J. W. Rationalizing membrane protein overexpression. *Trends Biotechnol* **24**, 364-371 (2006).
6. Junge, F. et al. Large-scale production of functional membrane proteins. *Cell Mol Life Sci* **65**, 1729-1755 (2008).
7. Miroux, B. & Walker, J. E. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**, 289-298 (1996).
8. Blois, T. M. & Bowie, J. U. G-protein-coupled receptor structures were not built in a day. *Protein Sci* **18**, 1335-1342 (2009).
9. Maeda, S. & Schertler, G. F. Production of GPCR and GPCR complexes for structure determination. *Curr Opin Struct Biol* **23**, 381-392 (2013).
10. Hartley, J. L. Cloning technologies for protein expression and purification. *Curr Opin Biotechnol* **17**, 359-366 (2006).
11. Hunt, I. From gene to protein: a review of new and enabling technologies for multi-

parallel protein expression. *Protein Expr Purif* **40**, 1-22 (2005).
12. Koehn, J. & Hunt, I. High-Throughput Protein Production (HTPP): a review of enabling technologies to expedite protein production. *Methods Mol Biol* **498**, 1-18 (2009).
13. Lesley, S. A. Parallel methods for expression and purification. *Methods Enzymol* **463**, 767-785 (2009).
14. Walhout, A. J. et al. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol* **328**, 575-592 (2000).
15. Colwill, K. et al. Modification of the Creator recombination system for proteomics applications--improved expression by addition of splice sites. *BMC Biotechnol* **6**, 13 (2006).
16. Geertsma, E. R. & Dutzler, R. A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry* **50**, 3272-3278 (2011).
17. Lundstrom, K. et al. Structural genomics on membrane proteins: comparison of more than 100 GPCRs in 3 expression systems. *J Struct Funct Genomics* **7**, 77-91 (2006).
18. Krettler, C., Reinhart, C. & Bevans, C. G. Expression of GPCRs in Pichia pastoris for structural studies. *Methods Enzymol* **520**, 1-29 (2013).
19. Freigassner, M., Pichler, H. & Glieder, A. Tuning microbial hosts for membrane protein production. *Microb Cell Fact* **8**, 69 (2009).
20. Sorensen, H. P. Towards universal systems for recombinant gene expression. *Microb Cell Fact* **9**, 27 (2010).
21. Panneels, V., Kock, I., Krijnse-Locker, J., Rezgaoui, M. & Sinning, I. Drosophila photoreceptor cells exploited for the production of eukaryotic membrane proteins: receptors, transporters and channels. *PLoS One* **6**, e18478 (2011).
22. Mizutani, K., Yoshioka, S., Mizutani, Y., Iwata, S. & Mikami, B. High-throughput construction of expression system using yeast Pichia pastoris, and its application to membrane proteins. *Protein Expr Purif* **77**, 1-8 (2011).
23. Sahdev, S., Khattar, S. K. & Saini, K. S. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem* **307**, 249-264 (2008).
24. Terpe, K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **72**, 211-222 (2006).
25. Schlegel, S. et al. Optimizing membrane protein overexpression in the Escherichia coli strain Lemo21(DE3). *J Mol Biol* **423**, 648-659 (2012).
26. Drew, D., Lerch, M., Kunji, E., Slotboom, D. J. & de Gier, J. W. Optimization of membrane protein overexpression and purification using GFP fusions. *Nat Methods* **3**, 303-313 (2006).
27. Roosild, T. P. et al. NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. *Science* **307**, 1317-1321 (2005).
28. Dvir, H. & Choe, S. Bacterial expression of a eukaryotic membrane protein in fusion to various Mistic orthologs. *Protein Expr Purif* **68**, 28-33 (2009).
29. Kefala, G., Kwiatkowski, W., Esquivies, L., Maslennikov, I. & Choe, S. Application of Mistic to improving the expression and membrane integration of histidine kinase receptors from Escherichia coli. *J Struct Funct Genomics* **8**, 167-

172 (2007).
30. Blain, K. Y., Kwiatkowski, W. & Choe, S. The functionally active Mistic-fused histidine kinase receptor, EnvZ. *Biochemistry* **49**, 9089-9095 (2010).
31. Petrovskaya, L. E. et al. Expression of G-protein coupled receptors in Escherichia coli for structural studies. *Biochemistry (Mosc)* **75**, 881-891 (2010).
32. Midgett, C. R. & Madden, D. R. Breaking the bottleneck: eukaryotic membrane protein expression for high-resolution structural studies. *J Struct Biol* **160**, 265-274 (2007).
33. Mierau, I. & Kleerebezem, M. 10 years of the nisin-controlled gene expression system (NICE) in Lactococcus lactis. *Appl Microbiol Biotechnol* **68**, 705-717 (2005).
34. Kunji, E. R. et al. Eukaryotic membrane protein overproduction in Lactococcus lactis. *Curr Opin Biotechnol* **16**, 546-551 (2005).
35. Zhou, X. X., Li, W. F., Ma, G. X. & Pan, Y. J. The nisin-controlled gene expression system: construction, application and improvements. *Biotechnol Adv* **24**, 285-295 (2006).
36. Frelet-Barrand, A. et al. Lactococcus lactis, an alternative system for functional expression of peripheral and intrinsic Arabidopsis membrane proteins. *PLoS One* **5**, e8746 (2010).
37. Linares, D. M., Geertsma, E. R. & Poolman, B. Evolved Lactococcus lactis strains for enhanced expression of recombinant membrane proteins. *J Mol Biol* **401**, 45-55 (2010).
38. Kiley, P. J. & Kaplan, S. Molecular genetics of photosynthetic membrane biosynthesis in Rhodobacter sphaeroides. *Microbiol Rev* **52**, 50-69 (1988).
39. Laible, P. D., Scott, H. N., Henry, L. & Hanson, D. K. Towards higher-throughput membrane protein production for structural genomics initiatives. *J Struct Funct Genomics* **5**, 167-172 (2004).
40. Szamecz, B., Urban, G., Rubiera, R., Kucsera, J. & Dorgai, L. Identification of four alcohol oxidases from methylotrophic yeasts. *Yeast* **22**, 669-676 (2005).
41. Andre, N. et al. Enhancing functional production of G protein-coupled receptors in Pichia pastoris to levels required for structural studies via a single expression screen. *Protein Sci* **15**, 1115-1126 (2006).
42. King, K., Dohlman, H. G., Thorner, J., Caron, M. G. & Lefkowitz, R. J. Control of yeast mating signal transduction by a mammalian beta 2-adrenergic receptor and Gs alpha subunit. *Science* **250**, 121-123 (1990).
43. Payette, P., Gossard, F., Whiteway, M. & Dennis, M. Expression and pharmacological characterization of the human M1 muscarinic receptor in Saccharomyces cerevisiae. *FEBS Lett* **266**, 21-25 (1990).
44. Sizmann, D. et al. Production of adrenergic receptors in yeast. *Receptors Channels* **4**, 197-203 (1996).
45. Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem* **73**, 1019-1049 (2004).
46. Hamilton, S. R. et al. Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* **313**, 1441-1443 (2006).
47. Lee, A. G. How lipids affect the activities of integral membrane proteins. *Biochim Biophys Acta* **1666**, 62-87 (2004).
48. Tate, C. G. et al. Comparison of seven different heterologous protein expression

systems for the production of the serotonin transporter. *Biochim Biophys Acta* **1610**, 141-153 (2003).
49. Condreay, J. P. & Kost, T. A. Baculovirus expression vectors for insect and mammalian cells. *Curr Drug Targets* **8**, 1126-1131 (2007).
50. Philipps, B., Rotmann, D., Wicki, M., Mayr, L. M. & Forstner, M. Time reduction and process optimization of the baculovirus expression system for more efficient recombinant protein production in insect cells. *Protein Expr Purif* **42**, 211-218 (2005).
51. Possee, R. D. Baculoviruses as expression vectors. *Curr Opin Biotechnol* **8**, 569-572 (1997).
52. Emery, V. C. Baculovirus expression vectors : choice of expression vector. *Methods Mol Biol* **8**, 287-307 (1992).
53. Jarvis, D. L. Baculovirus-insect cell expression systems. *Methods Enzymol* **463**, 191-222 (2009).
54. Lundstrom, K. Semliki Forest virus vectors for rapid and high-level expression of integral membrane proteins. *Biochim Biophys Acta* **1610**, 90-96 (2003).
55. Raunser, S., Haase, W., Bostina, M., Parcej, D. N. & Kuhlbrandt, W. High-yield expression, reconstitution and structure of the recombinant, fully functional glutamate transporter GLT-1 from Rattus norvegicus. *J Mol Biol* **351**, 598-613 (2005).
56. Derouazi, M. et al. Serum-free large-scale transient transfection of CHO cells. *Biotechnol Bioeng* **87**, 537-545 (2004).
57. Geisse, S., Jordan, M. & Wurm, F. M. Large-scale transient expression of therapeutic proteins in mammalian cells. *Methods Mol Biol* **308**, 87-98 (2005).
58. Baldi, L., Hacker, D. L., Adam, M. & Wurm, F. M. Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. *Biotechnol Lett* **29**, 677-684 (2007).
59. Aricescu, A. R. et al. Eukaryotic expression: developments for structural proteomics. *Acta Crystallogr D Biol Crystallogr* **62**, 1114-1124 (2006).
60. Eifler, N. et al. Functional expression of mammalian receptors and membrane channels in different cells. *J Struct Biol* **159**, 179-193 (2007).
61. Hofmann, C. et al. Efficient gene transfer into human hepatocytes by baculovirus vectors. *Proc Natl Acad Sci U S A* **92**, 10099-10103 (1995).
62. Shoji, I. et al. Efficient gene transfer into various mammalian cells, including non-hepatic cells, by baculovirus vectors. *J Gen Virol* **78**, 2657-2664 (1997).
63. Condreay, J. P., Witherspoon, S. M., Clay, W. C. & Kost, T. A. Transient and stable gene expression in mammalian cells transduced with a recombinant baculovirus vector. *Proc Natl Acad Sci U S A* **96**, 127-132 (1999).
64. Ames, R. et al. BacMam recombinant baculoviruses in G protein-coupled receptor drug discovery. *Receptors Channels* **10**, 99-107 (2004).
65. Kost, T. A., Condreay, J. P. & Jarvis, D. L. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat Biotechnol* **23**, 567-575 (2005).
66. Hassan, N. J., Pountney, D. J., Ellis, C. & Mossakowska, D. E. BacMam recombinant baculovirus in transporter expression: a study of BCRP and OATP1B1. *Protein Expr Purif* **47**, 591-598 (2006).
67. Scott, M. J. et al. Efficient expression of secreted proteases via recombinant

BacMam virus. *Protein Expr Purif* **52**, 104-116 (2007).
68. Neddermann, P. et al. A novel, inducible, eukaryotic gene expression system based on the quorum-sensing transcription factor TraR. *EMBO Rep* **4**, 159-165 (2003).
69. Lipscomb, M. L., Mowry, M. C. & Kompala, D. S. Production of a secreted glycoprotein from an inducible promoter system in a perfusion bioreactor. *Biotechnol Prog* **20**, 1402-1407 (2004).
70. Ivanova, L., Brandli, J., Saudan, P. & Bachmann, M. F. Hybrid Sindbis/Epstein-Barr virus episomal expression vector for inducible production of proteins. *Biotechniques* **39**, 209-212 (2005).
71. Bach, M. et al. Fast set-up of doxycycline-inducible protein expression in human cell lines with a single plasmid based on Epstein-Barr virus replication and the simple tetracycline repressor. *FEBS J* **274**, 783-790 (2007).
72. Hiroaki, Y. et al. Implications of the aquaporin-4 structure on array formation and cell adhesion. *J Mol Biol* **355**, 628-639 (2006).
73. Reeves, P. J., Thurmond, R. L. & Khorana, H. G. Structure and function in rhodopsin: high level expression of a synthetic bovine opsin gene and its mutants in stable mammalian cell lines. *Proc Natl Acad Sci U S A* **93**, 11487-11492 (1996).
74. Mollaaghababa, R., Davidson, F. F., Kaiser, C. & Khorana, H. G. Structure and function in rhodopsin: expression of functional mammalian opsin in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **93**, 11482-11486 (1996).
75. Janssen, J. J., Bovee-Geurts, P. H., Merkx, M. & DeGrip, W. J. Histidine tagging both allows convenient single-step purification of bovine rhodopsin and exerts ionic strength-dependent effects on its photochemistry. *J Biol Chem* **270**, 11222-11229 (1995).
76. Reeves, P. J., Thurmond, R. L. & Khorana, H. G. Structure and function in rhodopsin: high level expression of a synthetic bovine opsin gene and its mutants in stable mammalian cell lines. *Proc Natl Acad Sci U S A* **93**, 11487-11492 (1996).
77. Standfuss, J. et al. Crystal structure of a thermally stable rhodopsin mutant. *J Mol Biol* **372**, 1179-1188 (2007).
78. Berrow, N. S. et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res* **35**, e45 (2007).
79. Zubay, G. In vitro synthesis of protein in microbial systems. *Annu Rev Genet* **7**, 267-287 (1973).
80. Kigawa, T. et al. Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett* **442**, 15-19 (1999).
81. Kigawa, T. et al. Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. *J Struct Funct Genomics* **5**, 63-68 (2004).
82. Kim, T. W., Kim, D. M. & Choi, C. Y. Rapid production of milligram quantities of proteins in a batch cell-free protein synthesis system. *J Biotechnol* **124**, 373-380 (2006).
83. Klammt, C. et al. High level cell-free expression and specific labeling of integral membrane proteins. *Eur J Biochem* **271**, 568-580 (2004).
84. Wuu, J. J. & Swartz, J. R. High yield cell-free production of integral membrane proteins without refolding or detergents. *Biochim Biophys Acta* **1778**, 1237-1250 (2008).
85. Klammt, C. et al. Facile backbone structure determination of human membrane

proteins by NMR spectroscopy. *Nat Methods* **9**, 834-839 (2012).
86. Kohno, T. & Endo, Y. Production of protein for nuclear magnetic resonance study using the wheat germ cell-free system. *Methods Mol Biol* **375**, 257-272 (2007).
87. Vinarov, D. A., Loushin Newman, C. L. & Markley, J. L. Wheat germ cell-free platform for eukaryotic protein production. *FEBS J* **273**, 4160-4169 (2006).
88. Katzen, F. et al. Insertion of membrane proteins into discoidal membranes using a cell-free protein expression approach. *J Proteome Res* **7**, 3535-3542 (2008).
89. Kalmbach, R. et al. Functional cell-free synthesis of a seven helix membrane protein: in situ insertion of bacteriorhodopsin into liposomes. *J Mol Biol* **371**, 639-648 (2007).
90. Klammt, C. et al. Evaluation of detergents for the soluble expression of alpha-helical and beta-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *FEBS J* **272**, 6024-6038 (2005).
91. Klammt, C. et al. Polymer-based cell-free expression of ligand-binding family B G-protein coupled receptors without detergents. *Protein Sci* **20**, 1030-1041 (2011).
92. Junge, F. et al. Modulation of G-protein coupled receptor sample quality by modified cell-free expression protocols: a case study of the human endothelin A receptor. *J Struct Biol* **172**, 94-106 (2010).
93. Maslennikov, I. et al. Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis. *Proc Natl Acad Sci U S A* **107**, 10902-10907 (2010).
94. Roos, C. et al. Co-translational association of cell-free expressed membrane proteins with supplied lipid bilayers. *Mol Membr Biol* **30**, 75-89 (2013).
95. Etzkorn, M. et al. Cell-free expressed bacteriorhodopsin in different soluble membrane mimetics: biophysical properties and NMR accessibility. *Structure* **21**, 394-401 (2013).
96. Lee, A. G. Biological membranes: the importance of molecular detail. *Trends Biochem Sci* **36**, 493-500 (2011).
97. Soubias, O. & Gawrisch, K. The role of the lipid matrix for structure and function of the GPCR rhodopsin. *Biochim Biophys Acta* **1818**, 234-240 (2012).
98. Opekarova, M. & Tanner, W. Specific lipid requirements of membrane proteins--a putative bottleneck in heterologous expression. *Biochim Biophys Acta* **1610**, 11-22 (2003).
99. Spector, A. A. & Yorek, M. A. Membrane lipid composition and cellular function. *J Lipid Res* **26**, 1015-1035 (1985).
100. Serebryany, E., Zhu, G. A. & Yan, E. C. Artificial membrane-like environments for in vitro studies of purified G-protein coupled receptors. *Biochim Biophys Acta* **1818**, 225-233 (2012).
101. Grisshammer, R. Purification of recombinant G-protein-coupled receptors. *Methods Enzymol* **463**, 631-645 (2009).
102. Chiu, M. L., Tsang, C., Grihalde, N. & MacWilliams, M. P. Over-expression, solubilization, and purification of G protein-coupled receptors for structural biology. *Comb Chem High Throughput Screen* **11**, 439-462 (2008).
103. Jackson, M. L. & Litman, B. J. Rhodopsin-phospholipid reconstitution by dialysis removal of octyl glucoside. *Biochemistry* **21**, 5601-5608 (1982).
104. Franke, R. R., Sakmar, T. P., Graham, R. M. & Khorana, H. G. Structure and function in rhodopsin. Studies of the interaction between the rhodopsin cytoplasmic

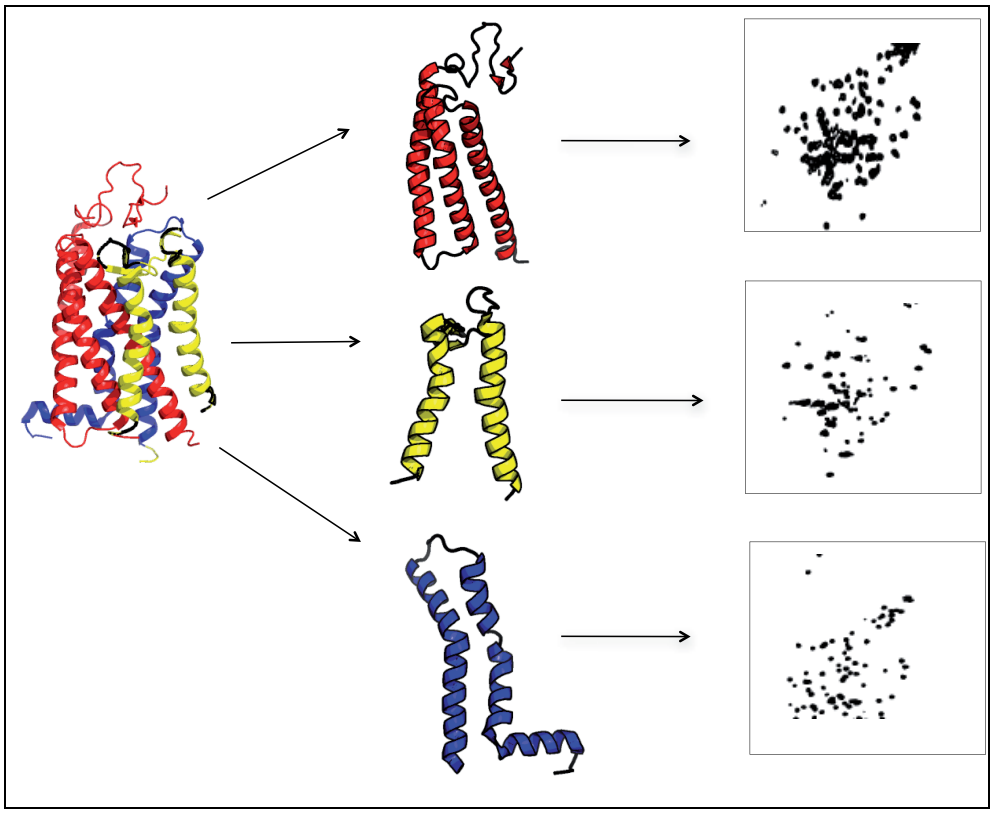
domain and transducin. *J Biol Chem* **267**, 14767-14774 (1992).
105. Okada, T. et al. X-Ray diffraction analysis of three-dimensional crystals of bovine rhodopsin obtained from mixed micelles. *J Struct Biol* **130**, 73-80 (2000).
106. Palczewski, K. et al. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-745 (2000).
107. Glover, K. J. et al. Structural evaluation of phospholipid bicelles for solution-state studies of membrane-associated biomolecules. *Biophys J* **81**, 2163-2171 (2001).
108. Prosser, R. S., Hwang, J. S. & Vold, R. R. Magnetically aligned phospholipid bilayers with positive ordering: a new model membrane system. *Biophys J* **74**, 2405-2418 (1998).
109. Sanders, C. R. & Prosser, R. S. Bicelles: a model membrane system for all seasons? *Structure* **6**, 1227-1234 (1998).
110. Reeves, P. J., Hwa, J. & Khorana, H. G. Structure and function in rhodopsin: kinetic studies of retinal binding to purified opsin mutants in defined phospholipid-detergent mixtures serve as probes of the retinal binding pocket. *Proc Natl Acad Sci U S A* **96**, 1927-1931 (1999).
111. Park, S. H. et al. High-resolution NMR spectroscopy of a GPCR in aligned bicelles. *J Am Chem Soc* **128**, 7402-7403 (2006).
112. Schmidt, P. et al. A reconstitution protocol for the in vitro folded human G protein-coupled Y2 receptor into lipid environment. *Biophys Chem* **150**, 29-36 (2010).
113. Brandt, D. R., Asano, T., Pedersen, S. E. & Ross, E. M. Reconstitution of catecholamine-stimulated guanosinetriphosphatase activity. *Biochemistry* **22**, 4357-4362 (1983).
114. Luca, S. et al. The conformation of neurotensin bound to its G protein-coupled receptor. *Proc Natl Acad Sci U S A* **100**, 10706-10711 (2003).
115. Bayburt, T. H. & Sligar, S. G. Membrane protein assembly into Nanodiscs. *FEBS Lett* **584**, 1721-1727 (2010).
116. Ritchie, T. K. et al. Chapter 11 - Reconstitution of membrane proteins in phospholipid bilayer nanodiscs. *Methods Enzymol* **464**, 211-231 (2009).
117. Leitz, A. J., Bayburt, T. H., Barnakov, A. N., Springer, B. A. & Sligar, S. G. Functional reconstitution of Beta2-adrenergic receptors utilizing self-assembling Nanodisc technology. *Biotechniques* **40**, 601-2, 604, 606, passim (2006).
118. Yao, X. J. et al. The effect of ligand efficacy on the formation and stability of a GPCR-G protein complex. *Proc Natl Acad Sci U S A* **106**, 9501-9506 (2009).
119. Whorton, M. R. et al. A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc Natl Acad Sci U S A* **104**, 7682-7687 (2007).
120. Bayburt, T. H. et al. Monomeric rhodopsin is sufficient for normal rhodopsin kinase (GRK1) phosphorylation and arrestin-1 binding. *J Biol Chem* **286**, 1420-1428 (2011).
121. Tsukamoto, H., Sinha, A., DeWitt, M. & Farrens, D. L. Monomeric rhodopsin is the minimal functional unit required for arrestin binding. *J Mol Biol* **399**, 501-511 (2010).
122. Banerjee, S., Huber, T. & Sakmar, T. P. Rapid incorporation of functional rhodopsin into nanoscale apolipoprotein bound bilayer (NABB) particles. *J Mol Biol* **377**, 1067-1081 (2008).
123. Bayburt, T. H., Leitz, A. J., Xie, G., Oprian, D. D. & Sligar, S. G. Transducin

activation by nanoscale lipid bilayers containing one and two rhodopsins. *J Biol Chem* **282**, 14875-14881 (2007).
124. Knepp, A. M., Grunbeck, A., Banerjee, S., Sakmar, T. P. & Huber, T. Direct measurement of thermal stability of expressed CCR5 and stabilization by small molecule ligands. *Biochemistry* **50**, 502-511 (2011).
125. Jarvis, D. L. Baculovirus-insect cell expression systems. *Methods Enzymol* **463**, 191-222 (2009).
126. Hauser, H. Short-chain phospholipids as detergents. *Biochim Biophys Acta* **1508**, 164-181 (2000).
127. Krueger-Koplin, R. D. et al. An evaluation of detergents for NMR structural studies of membrane proteins. *J Biomol NMR* **28**, 43-57 (2004).
128. McGuire, A. M., Matsuo, H. & Wagner, G. Internal and overall motions of the translation factor eIF4E: cap binding and insertion in a CHAPS detergent micelle. *J Biomol NMR* **12**, 73-88 (1998).
129. Sanders, C. R. & Sonnichsen, F. Solution NMR of membrane proteins: practice and challenges. *Magn Reson Chem* **44 Spec No**, S24-S40 (2006).
130. Hopper, J. T. et al. Detergent-free mass spectrometry of membrane protein complexes. *Nat Methods* **10**, 1206-1208 (2013).
131. Roos, C. et al. High-level cell-free production of membrane proteins with nanodiscs. *Methods Mol Biol* **1118**, 109-130 (2014).
132. Raschle, T. et al. Structural and functional characterization of the integral membrane protein VDAC-1 in lipid bilayer nanodiscs. *J Am Chem Soc* **131**, 17777-17779 (2009).
133. Yu, T. Y., Raschle, T., Hiller, S. & Wagner, G. Solution NMR spectroscopic characterization of human VDAC-2 in detergent micelles and lipid bilayer nanodiscs. *Biochim Biophys Acta* **1818**, 1562-1569 (2012).
134. Zoonens, M., Catoire, L. J., Giusti, F. & Popot, J. L. NMR study of a membrane protein in detergent-free aqueous solution. *Proc Natl Acad Sci U S A* **102**, 8893-8898 (2005).

Chapter II: Biosynthesis and spectroscopic characterization of 2-TM fragments encompassing the sequence of a human GPCR, the Y4 receptor

1. Abstract

This paper presents a divide-and-conquer approach towards solution structures of G-protein coupled receptors. The human Y4 receptor is dissected into 2-3 transmembrane helix fragments, which are individually studied by solution NMR. We systematically compare various aspects of biosynthetic routes for expression of the fragments in *E. coli* and discuss purification strategies. In particular, we have compared the production of the TM fragments into inclusion-bodies using the ΔTrp leader sequence, with the membrane-directed expression using Mistic as the fusion partner, and developed methods for enzymatic cleavage. In addition, direct expression of 2-TM fragments into inclusion bodies is a successful route in some cases. With the exception of TM13, we could produce all fragments in isotope-labeled form in quantities sufficient for NMR studies.



Complete to nearly complete backbone resonance assignments are obtained for the first

two helices as well as for helix 5 and 7 and a high degree for those of TM6, while conformational exchange processes result in disappearance of many signals from TM4. In addition, complete assignments are obtained for all residues from the N-terminal domain, the extracellular as well as the cytosolic loops (with exception of a undecapeptide segment in EC2) and for the complete cytosolic C-terminal tail: In total backbone resonances of 78% of all residues have been assigned for the Y4 receptor. Predictions of secondary structure based on backbone chemical shifts indicate that most residues from the TM regions occupy helical conformations with exceptions around polar residues or Pro sites. However, the domain borders are often slightly different from what is predicted from homology models. We suggest that the obtained chemical shifts may be useful in assigning the full-length receptor.

2. Introduction:

Our knowledge of structural biology of G-protein coupled receptors has advanced significantly in the last decade. High-resolution structures determined from single crystal diffraction now are available for the ground states of a number of receptors, e.g. bovine rhodopsin,[1] the turkey β_1[2] and human β_2[3] receptors, the human adenosine A2A receptor,[4] the human dopamine D3 receptor,[5] the human CXCR4 receptor[6] and the human histamine H1 receptor.[7] Mostly, these structures were determined in the presence of inverse antagonists, or for mutants that display little basal activity. Moreover, recent structures of GPCRs allow deciphering the mechanism of activation. To this end structures of GPCR in the agonist-bound state[8-11] or for mutants with constitutive activity[12] have been determined. Moreover, the ternary complex of the β_2-receptor, its ligand and G-protein bound has been investigated.[13] In case of rhodopsin, the structure of its ground state,[1] its state when bound to all-trans retinal (metarhodopsin II) and in retinal-free opsin[14,15] allowed mapping the pathway of activation.

To allow crystallographic studies GPCRs had to be stabilized biochemically, either by inserting the sequence of T4 lysozyme into the third cytosolic loop, or by selecting for temperature-stabilized mutants (for a review see Bill *et al.*[16]). In principle, NMR spectroscopy is capable of studying structure and dynamics of proteins in solution. In

particular this is also the case for proteins that contain partially unstructured parts or longer loops, features that usually hamper crystallization. The high-quality structure of sensory rhodopsin[17,18] and the recently published NMR structure of proteorhodopsin,[19] although not true GPCRs, but also of other integral membrane proteins[20-22] indicate that structure determination is possible by solution NMR methods once spectra of sufficient quality can be recorded. While progress in this area can be expected, to our knowledge presently in the particular case of true GPCRs spectral quality is not sufficient to follow this approach. Moreover, biosynthesis, purification and refolding of isotope-labeled entire GPCRs are very challenging.

Herein we study large fragments of the human Y4 receptor, a class A GPCR, to learn about structural features of these entities, to develop methods for assigning entire GPCRs and to investigate folding of GPCRs. Previously, we determined the structure of a fragment corresponding to the first two transmembrane (TM) helices of the yeast Ste2p receptor.[23] That fragment could be expressed in good yields in all flavors of labeling, and was shown to adopt tertiary structure in detergent micelles. Moreover, we previously reported on the structure and ligand binding of the isolated extracellular domain of the Y4 receptor (N-Y4).[24] Furthermore, we described the biosynthesis and backbone assignment of the N-TM12 fragment.[25]

In this work we present a comprehensive overview on our efforts at studying a series of 2-3 TM fragments spanning the entire sequence of the Y4 receptor. We report on the biosynthesis of the fragments in *E. coli* comparing both direct expression methods as well as fusions that direct expression products into inclusion bodies or membranes. We present fingerprint spectra for the fragments, and demonstrate that backbone assignments are possible to a large extent. We discuss the results with respect to GPCR folding and speculate on the future use of the data for supporting assignments of entire receptors.

3. Results

3.1 Biosynthesis:

A biosynthetic method that proves useful for NMR studies needs to accomplish several tasks: expression yields should be such that an NMR sample (at least 200 but preferably 300μM in concentration) can be made from 1-2 L of culture. Perdeuteration is absolutely necessary in order to facilitate backbone assignments, unless a large set of selectively labeled proteins is made. The method must also be capable for producing the protein in a form from which it can be refolded, and from which formation of precipitates or high-molecular weight oligomers is unlikely.

Yields for direct expression of helical integral membrane proteins are often unsatisfyingly low. Several methods have been proposed to increase yields. The most popular method is the use of fusion proteins, and both soluble as well as insoluble fusions have been advertised. In case of entire GPCRs conceptually the best method clearly is direct expression into membranes, but the low yields essentially prevent this method from being useful for NMR studies. A promising strategy to direct expression into membranes is the use of Mistic, a membrane protein from *B. subtilis* that bypasses the translocon-mediated insertion into the membrane and therefore allows much higher yields.[26] An alternative strategy is clearly the direction into inclusion bodies followed by a refolding step (see Figure 1). The requirement for full deuteration is incompatible with most expression system, and essentially precludes the use of sf9 insect cells despite the fact that most of the GPCRs that were recently determined by crystallography were expressed in that host. To the best of our knowledge deuteration can only be achieved *in vivo* in *E. coli*,[27] and in the methylotrophic yeast *Pichia Pastoris*[28] or *in vitro* with cell-free expression methods.[29] In this work we have used *E. coli* for the reason of ease of use, and because comparably high expression levels can be achieved in this host.

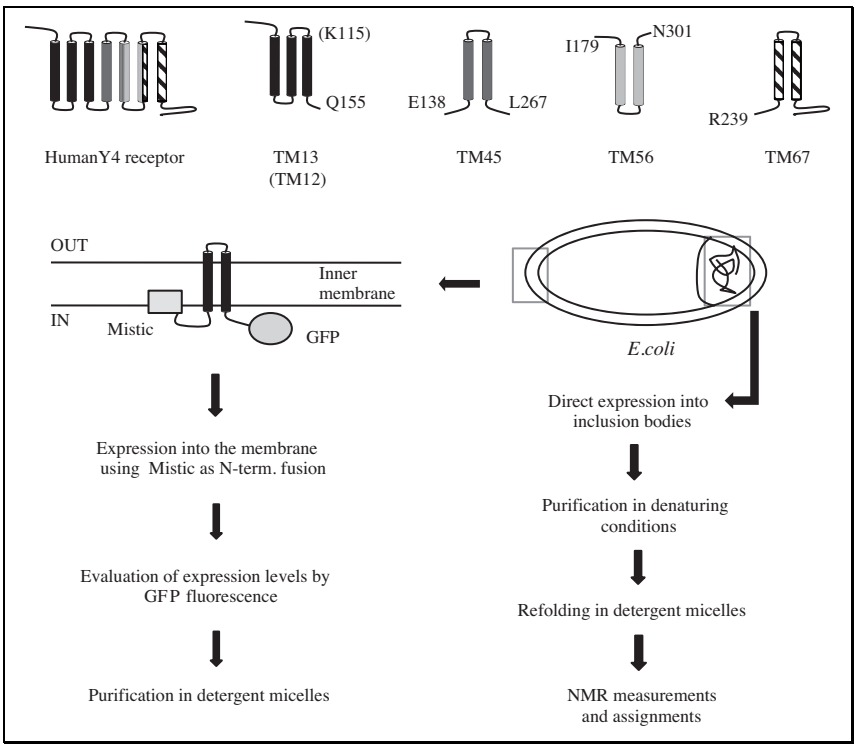


Figure 1: Overview of the constructs and the two strategies used for expression and purification.

With the aim of improving direct expression into inclusion bodies, we have compared the expression levels using the original cDNA coding for the human Y4 receptor and the same sequence devoid of rare codons and optimized for the GC content of *E. coli*. We did not observe any significant improvement in expression when using the codon-optimized sequence. The regions of the Y4 receptor used in this study are spanning residues M1-Q155 (here referred to as TM13), E138-L267 (TM45), I179-N301 (TM56) and R239 – I375 (TM67) (Sup. Mat. Figure S2). Direct expression into inclusion bodies was possible for N-TM13, N-TM45, TM45 and TM56 using minimal medium, although with variable yields. We could demonstrate that spectra derived from samples refolded after the Ni-NTA purification usually do not display good quality, which we attribute to the presence of residual lipids that result in micro-heterogeneity and conformational exchange[25]. In contrast, a second step of purification using a C4 reverse-phase HPLC step resulted in more homogeneous protein that displayed improved spectra. Final yields of purified protein per liter of culture differed, depending on the exact construct, and was observed to be in the range between 0.4 (N-TM13), 0.8 (TM45) and 0.7 (TM56) mg (for a gel of

the purified proteins see Figure 2). An advantage of this procedure is that at the end of the purification and following lyophilization the protein is available in powder form and can be simply dissolved in any detergent, thereby avoiding a detergent exchange step - a particular advantage when using deuterated detergents.

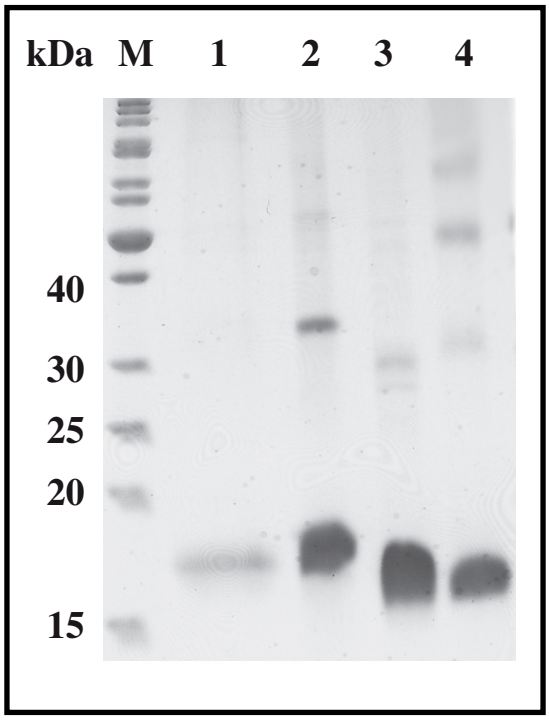


Figure 2: Coomassie-stained SDS-page of the fragments directly expressed into inclusion bodies and purified under denaturing conditions: M (marker), lane 1 TM13, lane 2 N- TM45, lane 3 TM45, lane 4 TM56.

In our experience, backbone assignments of membrane proteins of this size require highly deuterated proteins, in particular for experiments that relay magnetization onto the C_β carbons. Given the high content of some amino acids and the reduced signal dispersion of helical membrane proteins, knowledge of C_β chemical shifts is important for sequential assignments. Typically, expression yields drop by more than 50 % when growing cultures in deuterated water, even after extensive optimization. In our case TM56 and TM45 could not be expressed directly in perdeuterated form, while for the other constructs deuterated proteins sufficient for one NMR sample were obtained from 1 L of culture.

Some of the constructs could not be directly made, and therefore they were expressed as fusion proteins. Following earlier work of the Naider group we expressed TM67 as the fusion to the ΔTrp leader peptide (ΔTrp-LE). The ΔTrp-LE sequence is highly hydrophobic and results in accumulation of the expression product in inclusion bodies. In our experience this method is generic and usually results in reasonable quantities. An obvious problem with this (or most of the other fusions) is that the fusion partner must be removed. When working under denaturing conditions this can be achieved chemically using a CNBr cleavage step, with the limitation that only a single N-terminal Met residue is tolerated in the target protein. We also frequently encountered problems with the cleavage when Cys residues were present. Since the TM67 construct contains both Met as well as Cys residues we decided to utilize enzymatic cleavage as an alternative method. The particular problem with the latter is that the fusion protein must be soluble under conditions under which the protease is still active. We performed initial cleavage trials using the TEV or the 3C protease. Both enzymes are reported to be active in presence of low concentrations of denaturant or detergent. In a first approach we have solubilized the ΔTrp-TM67 fusion from inclusion bodies in a detergent buffer containing 1% dodecylphosphocholine (DPC). Although the purification using the IMAC was efficient and the 3C cleavage was almost 70%, we have experienced problems in completely removing cleaved ΔTrpLE and uncleaved ΔTrpLE-TM67 during the IMAC step. Although we have been able to produce a sample suitable for NMR using this method, the difficulties of detergent exchange, mostly due to the low CMC of DPC, precluded its further use. In order to increase cleavage efficiency, to improve solubility in denaturant at concentrations compatible with the enzymatic cleavage, and to improve binding to the Ni-NTA column, we have modified the original construct in the following way: i) a second deca-His tag was inserted prior to the 3C cleavage site. Since this tag is present in the part that is removed from the construct it does not cause further complications and even (slightly) increases solubility. ii) A 20 residue hydrophilic linker was inserted prior to the 3C site containing the sequence AKQFVEDNAENDEAEKLFNQ. The fusion protein was solubilized in 6M GdmCl, subjected to a Ni-NTA purification step and dialyzed against cleavage buffer in the presence of 3C protease. At sufficiently low concentrations of GdmCl the 3C protease refolds while the ΔTrp-TM67 fusion still

remains in solution. Under these conditions the cleavage was more than 80% efficient. Following dialysis, the precipitate material was redissolved in 6M GdmCl and was directly injected into the C4-HPLC. Using this methodology we were able to obtain 2 or 3 mg of purified protein from 1L of culture using M9 minimal medium in light or heavy water, respectively.

We noticed that yields for direct expression of N-TM12 were substantially higher than for the other 2-TM proteins. N-TM12 additionally contains the N-terminal 41-residue extracellular domain. The latter was investigated before in isolation, and low-affinity binding of the ligands was demonstrated.[24] We tested whether this small domain itself can serve as a fusion partner for some of the 2-TM proteins that we could not express directly, in particular for those not expressing in perdeuterated form. An attractive feature of this fusion is the very limited size, which even allows leaving it attached. Moreover, the fusion can be easily recognized because we assigned its resonances previously. We propose here to use this fusion to produce the perdeuterated form of a 2-TM protein that cannot be expressed in deuterated water directly, assign this protein, and then transfer the assignments from this protein onto the directly expressed. We have tested this approach for a smaller version of the construct TM45 (T147-H259), for which we failed to observe expression in deuterated water (data not shown). We were able to purify this construct in $^{15}N/^{13}C/^{2}H$ labeled form in sufficient yields (2 mg/liter of culture). Although we have not yet tested the use of the N-terminal domain systematically for the production of other truncated forms of the Y4 receptor, the use of N-terminal domain of the Y4 receptor represents an interesting alternative to the commonly used larger fusion proteins.

3.2 Expression and purification of Mistic fusions

All the fragments were initially tested for expression into the membrane of *E. coli* using Mistic fused at the N-terminus and the green-fluorescent protein (GFP) at the C-terminus. GFP fusions have proven valuable tools for evaluating expression levels of folded membrane proteins due to the fact the fluorescence levels are directly related to the amount of expressed protein.[30] All the fragments were well expressed into the membrane of *E. coli*, although the expression level of TM56 was lower (Figure 3). TM56 is the only fragment with an expected N-out-C-out topology (Figure 1) that should result

in a lower fluorescence signal due to misfolding of GFP in the periplasm. The total expression level of TM56 was confirmed to be lower from both an *in-gel* His-tag detection and a Coomassie-stained SDS-page gel (Figure 3). The expression of TM13 without any fusion directly into inclusion bodies resulted in low yields in comparison to other fragments. For this reason we have directed the expression of TM13 into the membrane of *E. coli* using the N-terminal Mistic fusion. Initial expression screenings revealed that Mistic-TM13 was better expressed in Rosetta pLysS cells at 16 °C, resulting in overall yields of the fusion construct after the IMAC purification in the range of 2 to 4 mg from 1 liter of LB culture. A solubilization assay[31] showed that the construct could not be extracted from the membranes using the detergent LDAO (Figure S1), a detergent that was suggested previously to efficiently solubilize Mistic on its own.[26] Other detergents suitable for NMR were tested, and DPC and LPPG proved to be most efficient. Finally, large scale purification and 3C protease cleavage were performed in presence of DPC, conditions under which the fusion was stable for several days at 4 °C. Following this procedure it was possible to purify 0.5 mg of target protein from 1 liter of LB culture and the final DPC concentration was ~ 15 mM as determined by NMR (Figure 4).

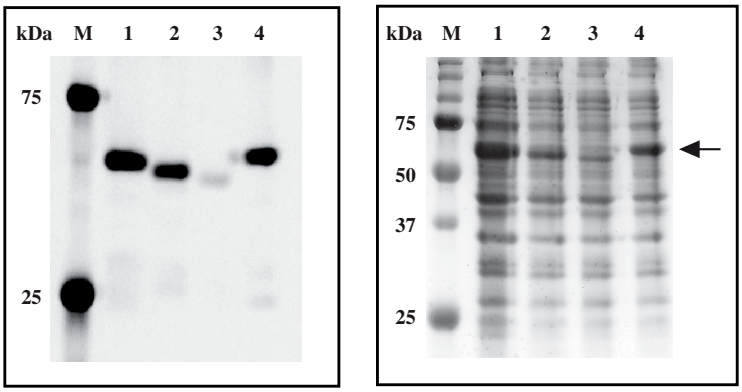


Figure 3: Expression of the fragments into the membrane of *E. coli* using Mistic at the N-terminus and GFP at the C-terminus. Left: in-gel fluorescence. Right: Coomassie stained SDS-page. On both gels lane 1 denotes Mistic-TM13-GFP, lane 2 Mistic-TM45-GFP, lane 3 Mistic-TM56-GFP and lane 4 Mistic-TM67-GFP.

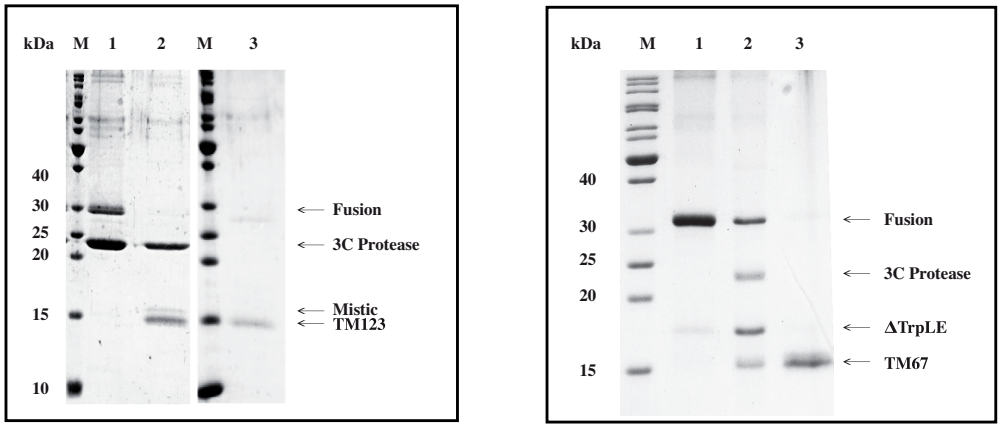


Figure 4: Cleavage and purification of Mistic–TM13 from membrane vesicles (left) and of ΔTrp-TM67 from GdmCl-solubilized inclusion bodies (right). Left: Lane 1: Mistic-TM13 before cleavage; lane 2: Mistic-TM13 after cleavage; lane 3: TM13 after reverse IMAC. Right: lane 1: ΔTrp-TM67 after IMAC, lane 2: after 3C cleavage, lane 3: after RP-HPLC.

3.3 Detergent screening and sample preparation

Spectra quality is of prime importance, and the extent to which resonances can be assigned very much depends on the latter. Detergent optimization therefore usually is the first step once a strategy for protein expression and purification has been established. Others and we are using $^{15}N,^{1}H$ correlation spectroscopy to evaluate the quality of spectra in a variety of detergents. The pros and cons of detergents have been extensively discussed in literature.[32-36] While certain classes of detergents have been advocated for use in solution NMR so far no particular detergent that in general is better than others has been identified, and detergents therefore need to be optimized for the protein in question. We have in our systems, however, observed that spectra in lyso-phosphocholine detergents generally behave better than detergents from the simple FOS series (e.g. DPC (FOS-12)). The group of Nietlispach observed superior spectra in C-7 DHPC.[18]

In the following we will describe the procedure used for TM67. Spectral quality in terms of signal dispersion and line-width as well as completeness of signals of TM67 was very good in 5% DHPC (C-7), but the sample started to precipitate after 4 hours of measurements at 47 °C. Other detergents screened were LMPC (70 peaks), LMPG (120

peaks) and LPPG (125 peaks) (see Supp. Mat. Figure S3). For diagnostic purposes we monitored the 7 glycine backbone and the 2 tryptophan sidechain peaks. Since all 7 peaks due to Gly residues were observed in LMPG and LPPG spectra with similar linewidths and peak intensities LPPG was selected as the detergent of choice since we were able to purchase the perdeuterated form of it. Representative spectra of ^{15}N labeled proteins of all investigated fragments are depicted in Figure 5.

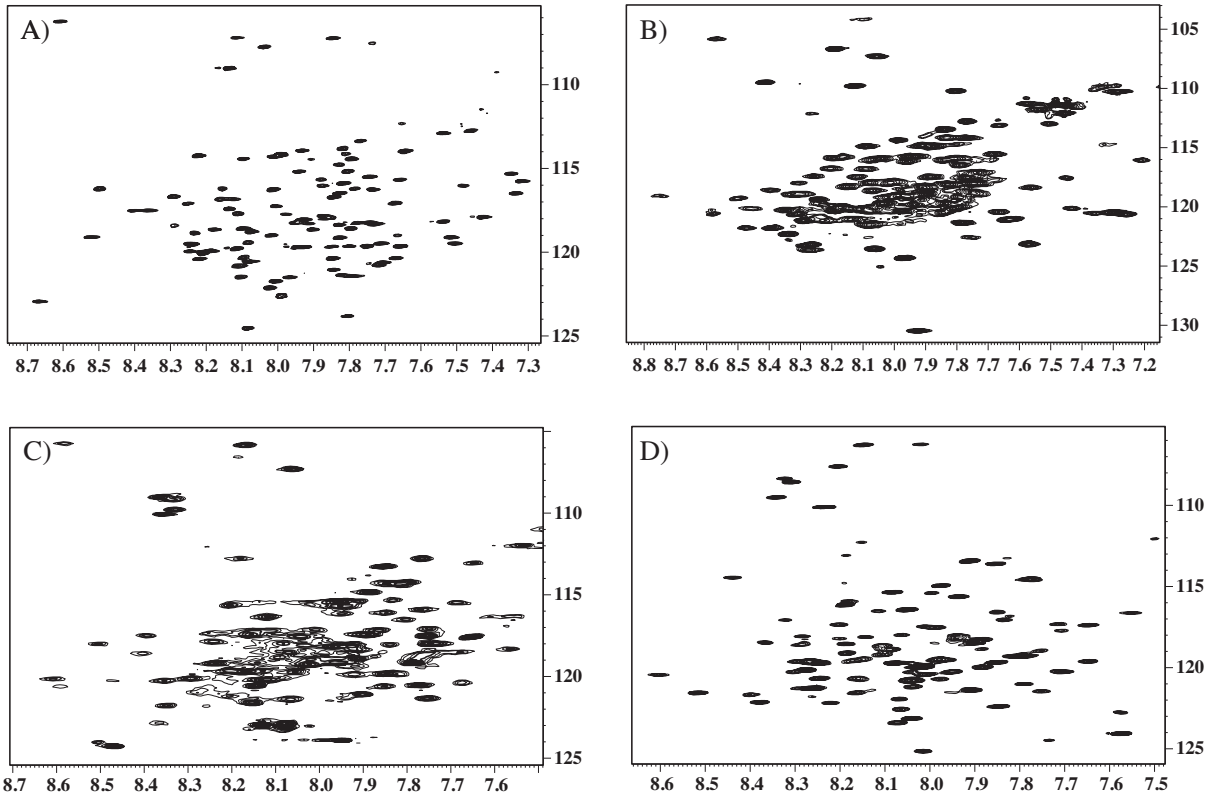


Figure 5: ^{15}N,^{1}H correlation spectra of TM1-TM2 (upper left), TM4-TM5 (upper right), TM5-TM6 (lower left) and TM6-TM7 (lower right). For annotation of cross peaks see the Supp. Mat.

3.4 Backbone Assignments

NMR spectroscopy of perdeuterated proteins is usually done with out-and-back type experiments.[37,38] Backbone assignments were determined with standard procedures by matching amide moieties via Cα and Cβ resonances in the HNCACB and HN(CO)CACB experiment, via CO frequencies in the HNCO and HN(CA)CO experiments, and via sequential amide protons NOEs in the [^{15}N,^{1}H]-NOESY experiments. The procedure used

for assignments in TM67 is briefly described in the following. Since the extent of peak overlap was unknown at the outset of the analysis we have started to adjust peak positions based on HNCO strips. The latter displayed good signal-to-noise peak for all peaks observed in the proton-nitrogen correlation map. Often correlations in HNCACB and HN(CO)CACB spectra were incomplete and in those cases HNCA and HN(CO)CA spectra were helpful to at least obtain common Cα frequencies. Assignments were straightforward for residues of segments that are not micelle-embedded as well as for residues of the putative TM7, resulting in unambiguous assignments for 24, 14 and 48 residues in the ICL3, ECL3 and C-terminus, respectively, as well as for 21 out of 24 residues of TM7. A representative plot of strips from the HNCACB spectrum for a segment of TM7 is displayed in Figure 6 (next page). Although, peaks could generally also be mostly observed for residues of the hydrophobic TM6, the repetitive occurrence of Val residues (e.g. VNVVLVVMVVA) did not allow connecting observed fragments unambiguously. Overall, possible assignments were observed for more than 90% of residues of TM67, and in 80% unambiguous annotations were made.

As TM45 could not be expressed directly in perdeuterated form, we decided to instead use the C-terminal fusion to NY4, the extracellular N-terminal domain of the Y4 receptor, which we could obtain from cultures grown in heavy water. Although a strategy similar to the one described above for TM67 was used, we encountered serious problems for assignments in TM4. While TM5 and the C-terminal tail could be completely assigned, only 2 out of 25 residues from TM4 were observed. In addition, large fractions from the sequence N-terminal to TM4 as well as from the loop connecting TM4 and TM5 were missing. Altogether unambiguous backbone assignments were possible only for 59% of the residues. For more details see the Supp. Mat. (shown in Tables S3-S6; for annotated spectra see Figures S4-S6) We would like to note here that we were able to completely assign all peaks with reasonable intensity. Accordingly, all peaks from missing residues are likely to be exchange-broadened. In addition we have also expressed [^{15}N^{13}C]-TM45, a truncated version in which the fusion to the N-terminal domain is removed. While this protein displays lower-quality triple-resonance spectra due to the lack of deuteration the quality of the ^{15}N-NOESY was much better.

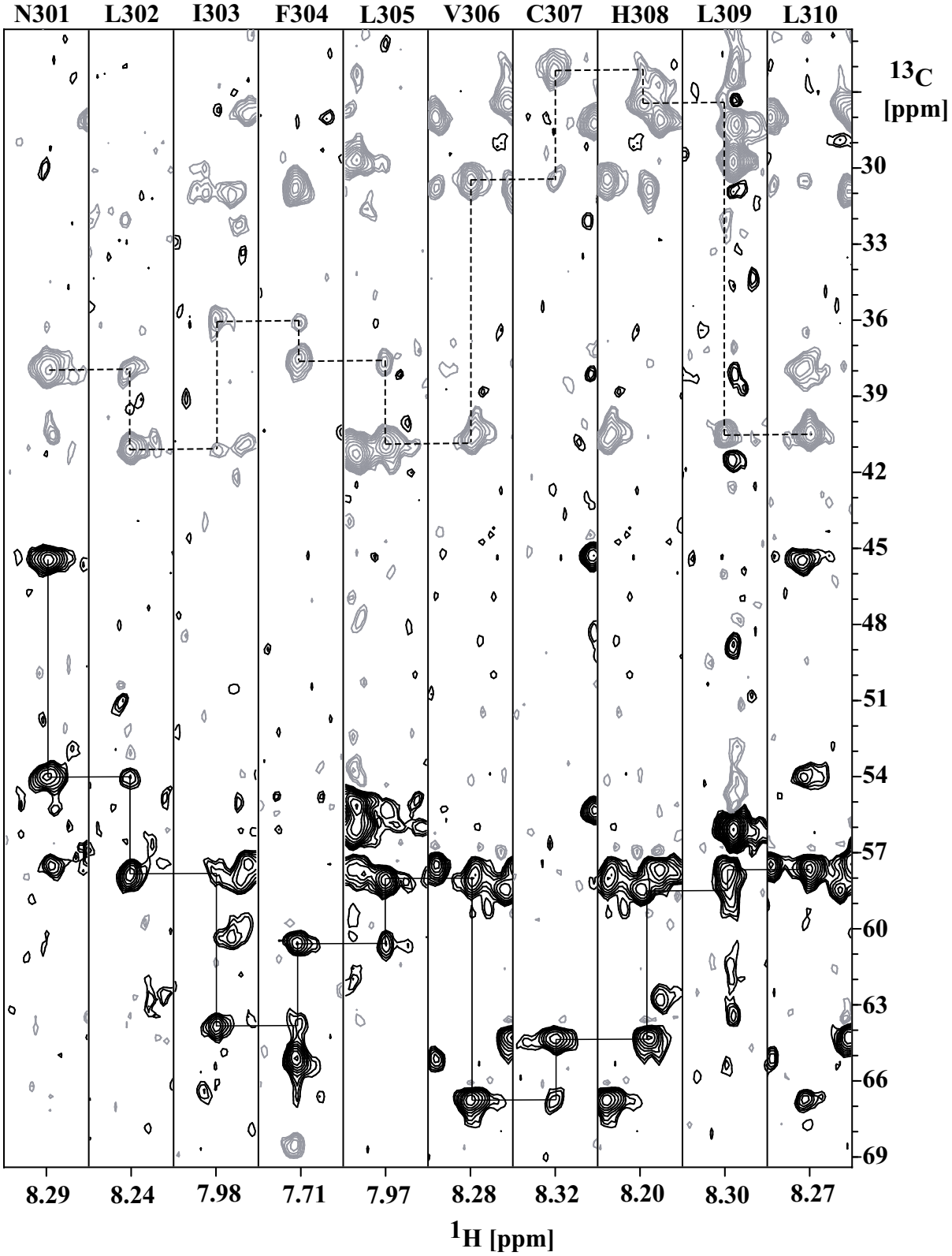

N301
L302
I303
F304
L305
V306
C307
H308
L309
L310
13C [ppm]
30
33
36
39
42
45
48
51
54
57
60
63
66
69
8.29
8.24
7.98
7.71
7.97
8.28
8.32
8.20
8.30
8.27
1H [ppm]

Figure 6: Plot displaying strips from the HNCACB spectrum of TM67 in a segment encompassing residues Asn-301 to Leu-310, that is part of TM7.

Most peaks from the TM segments did superimpose very well and peak positions could be adjusted using [^{15}N,^{1}H]-NOESY strips or common Cα, Cβ and C' shifts from the triple-resonance spectra. For this construct 75% of all backbone resonances were assigned. In particular further assignments in the putative TM4 and E2 loop were obtained (shown in Table S4).

Obviously, many peaks from residues in the TM segments were weak or missing. We speculate that this is due to conformational exchange processes. To gain insight into the details of this process we decided to measure a second data set for the protein N-TM45 in trifluoroethanol (TFE):water. This mixture has recently been used by the Naider lab to determine the structure of the TM12 fragment of the Ste2p GPCR.[39] Although tertiary structure is likely absent in this solvent mixture we hoped that these spectra might be useful to discover the missing signals. Due to the high quality of the triple-resonance spectra in this solvent we could rapidly obtain 90% backbone and side assignments. Nearly complete assignments were possible except for residues of the E2 loop segment and three residues from TM5 (see Supp. Mat. Table S5 and Figure S7), while gaps remained in the E2 loop. [^{15}N,^{1}H]-HSQC spectra of N-TM45 in LPPG detergent and in TFE/water are rather different, indicating that tertiary structure is altered in these two environments (see Figure S7). We are presently investigating whether assignments can be transferred from TFE:H_2O to LPPG spectra using comparisons of strips from ^{15}N-resolved NOESY spectra.

To investigate whether the presence of certain residues in central parts of TM4 is responsible for the observed line-broadening a number of point mutants such as W164A/ W164F or S171A/S171C was produced in ^{15}N-labeled form and evaluated based on the quality of their ^{15}N,^{1}H-correlation spectra. We reasoned that the polar Ser residue, for which no interaction partner exists in this fragment, may introduce flexibility in TM4.[40,41] Trp164 was further chosen for mutagenesis since Trp residues prefer to be

located in the interface rather than in the center.[42] Moreover, different Trp sidechain rotamers may induce a large effect on chemical shifts. We monitored signal intensity of Gly-159, that is clearly visible as a distinct single peak, and whose assignment can be readily adapted from the spectrum of the wild-type protein. In the wild-type protein this signal is significantly broadened indicating the presence of more than one state. In both the Ala and the Phe mutant Gly-159 is observed as a rather sharp peak with higher intensity (see Figure S8). While this mutant study is far from being complete and will be continued in our lab it indicates that much signal intensity for residues of TM4 may be recovered when a few residues are exchanged.

Backbone assignments of N-TM12 have been reported by us previously.[25] The triple-resonance spectra in that case were generally of much higher quality than for any other fragment described in this report, and almost complete backbone assignments could be obtained. In total backbone assignments for 281 out of 369 Y4 residues are made accounting for 78% of all residues (see also Figure 7).

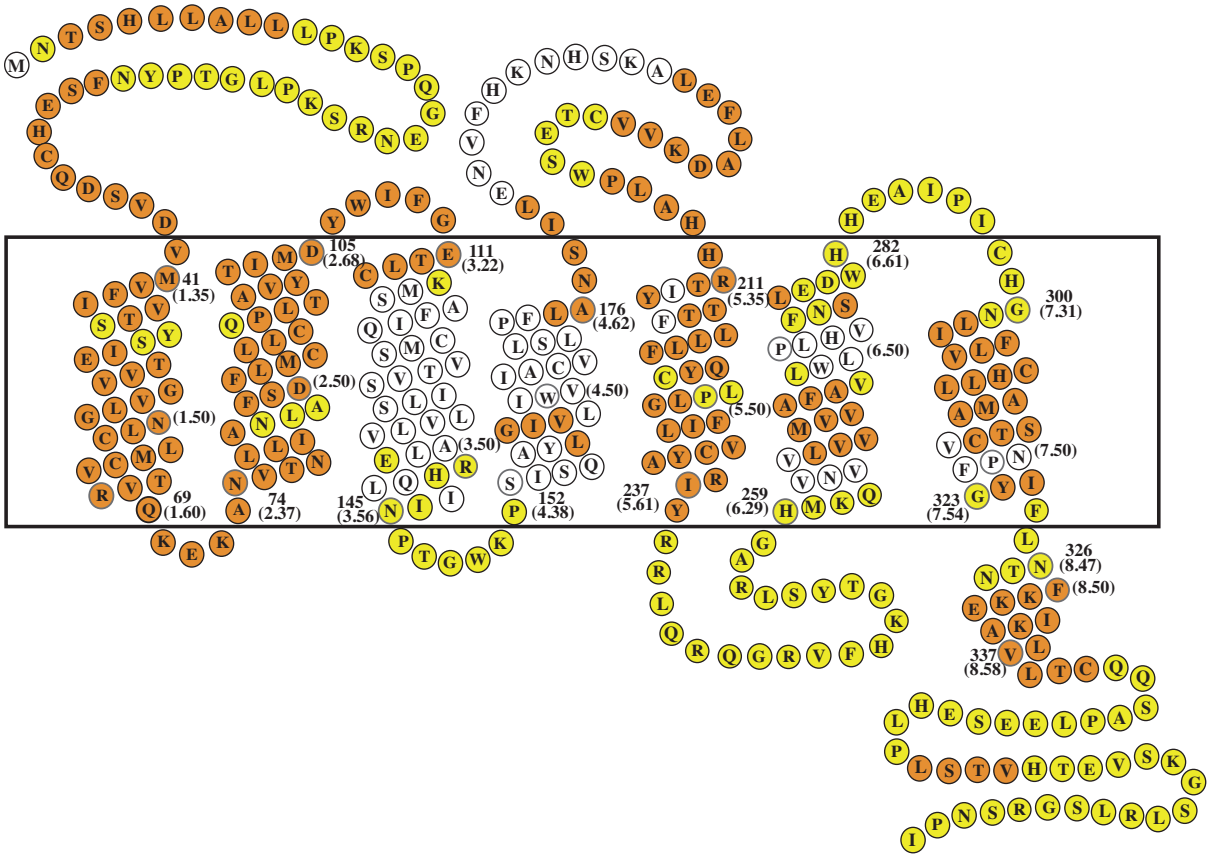


Figure 7: Snake plot of the Y4 receptor displaying secondary structure predictions based on backbone chemical shifts, as well as the present status of backbone assignments. Residues predicted to adopt helical or no secondary structure are displayed as orange or yellow spheres, respectively.

For a more detailed account of the status of assignments in all fragments see the Supp. Mat. Table S2. From the missing assignments 25 are due to residues from TM3, and hence from a fragment that we could not produce in deuterated from in sufficient amounts up to now.

4. Discussion

The intention of this work was to provide tools for structural studies of large fragments of GPCRs using solution NMR spectroscopy, with the aim of improving our knowledge on their folding and to obtain spectroscopic data that may help in tackling the entire receptors.

Structural studies of membrane proteins are usually hampered by low expression levels and by the difficulties in finding conditions under which the protein of interest is stable and displays spectra of sufficient quality.[43] The use of NMR spectroscopy needs additional prerequisites, such as that the requirement for isotope labeling at reasonable cost and for perdeuterated detergents as membrane mimicking environments. The presence of rare codons is one important factor that can reduce heterologous expression levels in bacteria.[44] This problem can be overcome using strains of *E. coli* that produce increased levels of rare tRNAs (Rosetta, Codon plus), or by synthesizing genes optimized for the *E. coli* codon usage. In our case, however, we did not observe any significant improvement in expression using the codon-optimized gene.

We have compared two methods for expressing and purifying the fragments in *E. coli*. The fragments were expressed in insoluble form by directing expression into inclusion bodies or in folded form by directing it into the inner membrane, where quantification of expression was greatly facilitated by the use of the C-terminal GFP tag.[45] Both methods

have been widely applied to membrane proteins and represent viable strategies for obtaining the target proteins in amounts compatible with structural studies.[35] The expression of a membrane protein into inclusion bodies usually requires refolding into a functional form, a step that involves careful optimization of the refolding buffer. Often only a small fraction of the target protein can be functionally refolded, especially in the case of alpha helical membrane proteins.[46-51] The large GPCR fragments used in this study are devoid of function, and we have therefore exploited the alternative route of directing them into inclusion bodies (IB). The possibility to purify the protein from these in absence of detergents allows to finally solubilize the lyophilized protein in the detergent of choice. This procedure thereby avoids difficult detergent exchange steps, and hence is particularly suitable when screening a larger number of detergents. Several proteins have been suggested as N-terminal fusions in order to enhance expression of GPCRs and direct them into IB[52]. In this study we used the ΔTrp-LE[53-55] for increasing the expression in IB for a fragment that could not be expressed directly. In addition, we presented a protocol for highly efficient cleavage from the ΔTrp-LE in detergent.

As an alternative expression method the fragments were directed into the membrane using the fusion to Mistic, and GFP added as a C-terminal fusion was exploited as an indicator for both expression level and protein folding. The presence of the GFP offers the additional advantage of directly detecting the target protein in standard SDS-page gels avoiding the use of immunoblot systems. The expression of TM13 into the inner membrane of *E. coli* with the help of Mistic represents a viable alternative to the expression into inclusion bodies. Presently, however, we prefer to use the IB directed methods because the yields are higher, and only a minimal amount of (expensive) detergent is required. Nevertheless, we suspect that the membrane-directed method may be useful for longer constructs that cannot be refolded with good efficiency.

A comparison of the expression levels of the various fragments described in this study reveals that it is difficult to know *a priori* which strategy will give the best results. Clearly, optimization of overexpression can be achieved only after extensive screening for several constructs and conditions. Provided that sufficient expression levels are

obtained direct expression into inclusion bodies is the method of choice. Unfortunately, protein quantities are often too low, and then the insoluble expression of suitable fusions into IBs is the more promising route to obtain these fragments. Whether biophysical properties of refolded protein and protein that is obtained from direct expression into the inner membrane is different in the case of fragments is subject to future studies.

The present study has demonstrated that a set of fragments comprising almost the complete sequence of the human Y4 GPCR can be obtained from expression in *E. coli* in isotope-labeled form. The corresponding $^{15}N,^{1}H$ correlation spectra display reasonable signal dispersion. Assignment of the backbone resonance was possible for 78% of all backbone resonances. However, we have also encountered serious problems with assignments of TM4 and TM6. While most of the problems in TM6 were related to repetitive sequences, no or weak peaks corresponding to residues of TM4 were observed in the spectra. We attribute this fact to conformational exchange processes in the detergent. Why exactly TM4 suffers so much from conformational exchange rather than other segments, in particular TM5 that is supposed to interact with TM4, is presently unclear to us and subject to further studies.

The knowledge of backbone chemical shifts can be used to rather reliably predict locations of secondary structure.[56,57] Figure 7 highlights the locations of predicted helical and loop regions in the secondary structure as predicted from the backbone chemical shifts using the program TALOS+, [56] and also summarizes the present state of backbone assignments. The data indicate reasonable overall agreement with the predictions. For both TM1 and TM2 the helical regions are longer than predicted from the homology model. We have previously determined the structure of the extracellular N-terminal domain, and observed the presence of a long rather flexible loop that is anchored on the micelle surface via helices at both termini.[24] A longer stretch with residues with α-helical propensities is also observed in the EC2 loop. Moreover, the presence of a 8th, cytosolic helix is in agreement with the predictions.

Significant destabilization of internal helices is observed within all TM segments, and mostly occurs around polar or Pro residues. Such polar residues are often located within the central TM segment and usually form contacts with polar residues from other helices. Whenever these other helices are not included in the fragment of interest the presence of these residues is expected to hamper proper integration of the TM helices into the membrane or micelle. When comparing the content of polar residues within central segments of TM helices between the 7 putative TM stretches it becomes clear that TM1, TM2, TM3 and TM7 form a polar core, whereas TM4, TM5 and TM6 are mostly devoid of polar residues. A systematic study of the location of termini in N- or C-terminally truncated fragments of the Y4 receptor has demonstrated that the C-terminal fragments better integrate into the membrane.[58] Moreover, we noticed in previous studies of the TM7 fragment from Ste2p that the TM7 was destabilized at positions of polar residues, which were positioned rather in the micelle-water interface than in the micelle interior.[59] Preliminary results on the structure determination of N-TM1-TM2 of the Y4 receptor indicate that the two helices are significantly destabilized around central polar residues (data not shown).

Recently, the structures of sensory rhodopsin[18] and of proteorhodopsin[19] have been determined by solution NMR methods. In both cases high-quality structures could be obtained, with very high degrees of backbone and considerable extent of sidechain assignments. In both cases assignments started from high-quality [^{15}N,^{1}H]-TROSY spectra, but both proteins despite their 7-TM bundle structure do not belong to the family of GPCRs since they are not coupled to G-proteins. Little is published on solution NMR of true GPCRs, but in the few available cases spectra are much more problematic displaying dramatically varying linewidths, and assignments were often only possible for residues from loops or the N- or C-terminal tails. We speculate that it is the inherent flexibility required for the mechanism of activation that results in the unfavorable spectroscopic properties that hamper their study by solution NMR. Crystallographers have conducted major efforts to stabilize either the ground or the activated state of these receptors,[60] and it may likely also present a viable route for NMR studies, and fragments may be useful to identify positions for modifications. Another interesting option of this approach is to obtain backbone and in particular side-chain assignments. We are

presently investigating to which extent such assignments can be adapted to spectra recorded on the entire receptor. Finally, studying fragments of GPCRs that contain a number (2-3) of entire TM helices therefore presents a possibility of studying folding of these proteins. Individual helices are believed to insert into and diffuse within the membrane until productive encounters between helices occur resulting in assembly of the tertiary structure.[61,62] We suggest that studying fragments may provide insight into the conformational preferences of the corresponding individual TM segments before the 7-TM bundle has been formed successfully.

5. Experimental Section

5.1 Enzymes and chemicals: Phusion DNA polymerase was purchased from Finnzymes. All the restriction enzymes were obtained from New England Biolabs, unless specified. Detergents dodecylphosphocholine (DPC), 1-palmitoyl-2-hydroxy-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (LPPG), 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine (DHPC), 1-myristoyl-2-hydroxy-*sn*-glycero-3-phosphocholine (LMPC) and 1-myristoyl-2-hydroxy-*sn*-glycero-3-phosphatidylglycerol (LMPG) were purchased from Anatrace (Maumee, OH) and Avanti (Alabaster, Alabama).

5.2 DNA techniques: All the cloning procedures were done following standard methods,[63] unless specified. The cDNA sequence of the human Y4 receptor was obtained from the Missouri S&T cDNA Resource Center, while the same sequence optimized for the *E. coli* codon usage was purchased from GeneArt. DNA sequences coding for the constructs TM13, TM45, N-TM45 and TM56 were amplified by PCR using as templates both human and the codon-optimized cDNA. All the forward and reverse primers used in this study are listed in Table S1 (Supp. Mat.). The PCR products were gel-purified and cloned into a pLC01 vector[55] between NdeI and BamHI restriction sites, and into a pET15b vector (Novagen) between NdeI and XhoI sites. Both plasmids are under the control of a T7 promoter and contain an N-terminal octa or deca-His tag, respectively. The extracellular N-terminal domain of the Y4 receptor (M1-M41), here used as a fusion protein for enhancing the expression of deuterated N-TM45, was

amplified by PCR and fused to the start of TM45 by overlapping PCR and cloned into a pLC01 vector between NdeI and BamHI sites. The DNA sequence coding for construct TM67 was amplified by PCR in order to include at the N-terminus an octa-His-tag and a 3C protease cleavage site (LEVLFQGP), separated by a hydrophilic 20 amino-acids linker (AKQFVEDNAENDEAEKLFNQ). The modified TM67 was then cloned into a pLC01 vector between HindIII and BamHI containing the N-terminal fusion TrpΔLE (a vector obtained from the Fred Naider Lab). The DNA sequence of Mistic was obtained from the Choe lab and was fused by overlapping PCR to the sequence coding for the cDNA of the Y4 receptor, introducing a 3C cleavage site between the two sequences. The PCR product was gel-purified and cloned into a pET21b vector between BamHI and XhoI sites, and was later on used as a template for the amplification of all the Mistic-constructs. For expression on large scale, the DNA sequence corresponding to Mistic-TM13 was amplified by PCR and cloned into a pET15b vector. DNA sequences corresponding to the constructs Mistic-TM13, Mistic-TM45, Mistic-TM56 and Mistic-TM67 were amplified by PCR using primers extended on their 5' and 3' by SapI sites using as templates vectors constructed previously[58] and cloned into a pBAD derivative vector (pBX3CGH) using the FX cloning method.[64] This vector is under the control of an arabinose promoter and contains the GFP gene following the cloning site.

Site-directed mutagenesis on the sequence coding for TM45 was done following the QuikChange® methodology (Stratagene). Primers for that purpose were designed using the QuikChange Primer Design program. Again, plasmids were sequenced and used to transform *E. coli* BL21-AI cells.

5.3 Strains and growth conditions: *E. coli* DH5α and *E. coli* BL21-AI (Invitrogen) were used as the host for the DNA manipulations and expression, respectively, for the constructs TM45 and its mutants as well as for TrpΔLE-TM67, while N-TM45 and TM56 were expressed in the *E. coli* strain C41.[65] Overnight cultures were grown in Luria-Bertani broth supplemented with ampicillin (100 μg/ml) at 37°C under vigorous shaking. Cells were then diluted 1:100 in 0.5 liter flasks (usually 2 L in total) of M9 medium supplemented with ampicillin (100 μg/ml) containing $^{15}NH_4Cl$ or $^{15}NH_4Cl/^{13}C$-

glucose as the sole nitrogen or carbon sources. For expression in deuterated water, cells were adapted to D_2O by plating them on a D_2O-M9 agar plate supplemented with ampicillin (100 μg/ml). Cells were grown at 37°C until the OD_{600} reached ~ 0.8 (TM45, N-TM45, TM56) or 0.6 (TrpΔLE-TM67), after which the temperature was lowered to 20°C. Expression was induced with arabinose (0.2%) in the case of the *E. coli* BL21-AI strain or with Isopropyl-β-D-thiogalactopyranoside (IPTG, 1 mM) for the *E. coli* C41 strain. Induced cells were harvested after 16 hours of growth in M9/H_2O or 40 hours of growth in M9/D_2O. Due to higher expression levels, the *E. coli* Rosetta pLysS strain (Novagen) was chosen as the host for Mistic-TM13. Cells were grown overnight while shaking at 37 °C, diluted to 1:1000 in LB broth supplemented with ampicillin (100 μg/ml), and grown at 37 °C until the OD_{600} was ~ 0.5. Subsequently, the temperature was lowered to 16 °C followed by induction with IPTG (0.5 mM). Finally induced cells were harvested after 24 hours. All the GFP fusions were expressed in *E. coli* SF100 [recA, Δlac, Δompt] (a gift from the Lolkema lab) in LB-broth (0.7 ml) using a 96 well plate (2 ml capacity). Cells from an overnight culture were diluted 1:100 and grown for 1 hour at 37 °C. After the temperature was lowered to 25 °C for 90 min, protein expression was induced with arabinose (0.2%) and the cells were harvested 4 hours later.

5.4 Purification from inclusion bodies: Cells were centrifuged at 6'000 g for 20 min. and the pellet was weighed and resuspended in a volume corresponding to ~ 4 times the weight of the cell pellet in Tris-buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl and 1 mM DTT). Cells were disrupted by sonication and inclusion bodies (IB) were separated by centrifugation at 10'000 g for 30 min. Inclusion bodies were solubilized in denaturation buffer (6 M GdmCl, 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM DTT) upon sonication (TM45, N-TM45, TM56) or by stirring at RT overnight (TrpΔLE-TM67). The insoluble material was removed by centrifugation (15'000 g for 30 min.)and the supernatant was mixed with the Ni-NTA resin (SIGMA) pre-equilibrated in the denaturation buffer, while stirring for 1 hour. The resin was washed with five column volumes of wash buffer (denaturation buffer containing 30 mM imidazole) and the target protein was eluted using a higher imidazole concentration (300 mM). The fractions corresponding to the elution peaks were pooled and dialyzed against 100 volumes of

water (overnight at RT) using a dialysis membrane (6-8 kDa, Spectrumlabs) to obtain the target protein in form of the precipitate. The precipitate was then solubilized in small quantities (usually 1 ml) of a highly reducing denaturation buffer (6 M GdmCl, 100 mM DTT, 250 mM β-mercaptoethanol (BME), 20 mM Tris-HCl, pH 8.0, 200 mM NaCl), after which the pH was lowered to ~ 2.0. The samples were injected into a C4 Reversed-Phase HPLC column (VYDAC) equilibrated with an aqueous solution containing trifluoracetic acid (TFA, 0.1%) and acetonitrile (ACN, 10%), operated at 55°C. The target proteins (TM56, TM45, TM13) were eluted at 68, 74, 90% of ACN, respectively, in a H_2O/ acetonitrile gradient.

The fusion construct TrpΔLE-TM67 was initially purified from inclusion bodies (IB) by solubilization in a detergent buffer (50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10 mM BME and 1% DPC) in a volume corresponding to 5 times the weight of the IB pellet. The solution was kept stirring for 12 hours at 4 ^{o}C and the insoluble fraction was removed by ultracentrifugation (40'000 g, 20 min.). The soluble fraction was then applied to a Ni-NTA column pre-equilibrated with detergent buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10 mM BME and 0.2% DPC). The column was washed and the target protein was eluted with the same buffer, additionally containing imidazole (30 mM and 300 mM respectively). Fractions corresponding to the elution peaks were pooled and the 3C protease was added at a final molar ratio of 1:5 of protease to target protein. The solution was then dialyzed overnight against 100 volumes of detergent buffer (50 mM Tris-HCl buffer pH 8.0, 0.2% DPC) and successively re-loaded onto the Ni-NTA column from which the flow through was collected. Unfortunately, the flow-through still contained most of the cleaved (His-tagged) TrpΔLE, which could be separated from the solution containing the fragment TM67 through multiple reloading on the Ni-NTA column. The same construct was also purified when solubilized inGdmCl (6 M), a procedure that was finally used for the preparation of the NMR samples. In this case, the GdmCl-solubilized fusion construct was purified from the inclusion bodies using the IMAC, and the fractions containing the fusion protein were pooled and mixed with the 3C protease in a final molar ratio of 1:5 (3C : TrpΔLE-TM67). The solution was subsequently dialyzed overnight against 100 volumes of protease cleavage buffer (50 mM Tris-HCl, pH 7.5, 200

mM NaCl and 10 mM BME). Upon removal of denaturant to sufficient levels, the 3C protease refolds, and cleaves the TrpΔLE fusion. Both fragments, TrpΔLE and TM67 precipitate, while most of the 3C protease remains in solution. The precipitate was then solubilized in 1 ml of denaturation buffer (6 M GdmCl, 100 mM DTT, 250 mM BME, 20 mM Tris-HCl pH 8.0, 200 mM NaCl), after which the pH was lowered to ~2.0. This solution was injected into the C4 RP-HPLC column, from which TM67 eluted at 61% acetonitrile using conditions described above. Fractions corresponding to the desired protein were lyophilized.

5.5 Purification of Mistic-TM13: Cells were harvested by centrifugation at 6'000 g for 15 min. at 4°C, and the cell pellet was resuspended in ice-cold KPi (50 mM), pH 7.5, NaCl (150 mM) such that the OD_{600} was ~150-200. After adding lysozyme (1 mg/ml final), DNase (20 μg/ml final) and $MgCl_2$ (1 mM final), cells were homogenized using a homogenizer device while keeping the suspension on ice and the solution was stirred for 1 hour at 4°C. Subsequently, cells were disrupted by three passes at 15kPsi (Emoulsiflex, cooled), and a protease inhibitor mixture (Roche) was added. After removal of unbroken cells by centrifugation at 15'000 g for 15 min. at 4 °C, the supernatant containing the membrane fraction was centrifuged at 40'000 g at 4 °C for 1 hour. The formed pellet was resuspended in KPi (50 mM), pH 7.5, NaCl (150 mM), glycerol (10%) using a Potter tube (Sigma) to a final concentration of 1 gr. of membrane/ 2 ml of buffer, and immediately frozen in liquid nitrogen. Membranes were solubilized by stirring for 1 hour at 4°C in KPi (50 mM, pH 7.5), NaCl (300 mM), glycerol (5%), DTT (1 mM) and DPC (1% wt). Insoluble material was then removed by centrifugation (40'000 g at 4°C for 1 hour), and the supernatant mixed with pre-equilibrated Ni-NTA resin. The column was washed with ten column volumes equilibration buffer containing imidazole (50 mM) and DPC (2 mM). The protein was eluted using a higher imidazole concentration (300 mM). The fractions corresponding to the desired protein were pooled together and dialyzed overnight against KPi (50 mM, pH 7.5), NaCl (300 mM), DTT (1 mM), glycerol (5%) and DPC (1 mM), adding 3C protease to the dialysis bag in a final molar ratio of 1:5 (protease: target). Following buffer optimization, best conditions were observed using KPi (50 mM), NaCl (300 mM), glycerol (2.5%), DPC (1 mM), DTT (1 mM), 4°C for 48

hours. The cleavage was done during the dialysis step that was necessary to remove imidazole from the Ni-NTA elution in the presence of DPC (1 mM). This low detergent concentration was chosen as a compromise for solubilizing the protein and retaining activity of the protease. After 48 hours, the cleavage was almost quantitative although part of the excised protein precipitated. The solution from the dialysis bag was then mixed with pre-equilibrated Ni-NTA resin, and the flow-through containing TM13 was collected and concentrated using an Amicon Centricon tube (Millipore, 30'000 MW cut-off).

5.6 Solubilization assay and in-gel fluorescence: Samples for the solubilization assay were prepared as described elsewhere.[31] Briefly, cell pellets corresponding to a total amount of proteins (2 mg) were resuspended in KPi (50 mM pH 7.5), NaCl (150 mM), glycerol (5%), a protease inhibitor mix (Roche) and glass beads (Sigma). Cells were disrupted by using the FastPrep device (Bio101), the supernatant containing the membranes was transferred to ultracentrifuge tubes and incubated with the detergents used in the assay (1% wt/vol of LDAO, DPC or LPPG). From these solutions small aliquots were taken, mixed with SDS-loading buffer, and stored until further use. These samples represented the total amount of protein before solubilization. Samples were incubated for 1 hour on ice in presence of detergents, and the insoluble material was removed by ultracentrifugation (100'000 g, 20 min.). Again, small aliquots from the supernatant were then mixed with the SDS-loading buffer. These samples represented the amount of solubilized protein (see Figure S1).

Aliquots were loaded on an acrylamide (15%) SDS-PAGE that was then blotted on a PVC membrane with the help of a semidry electroblotting system, followed by immunodetection (Roche). Chemiluminescence detection was performed using the Western light kit (Tropix, Inc.) and the Fujifilm LAS-3000 imaging system. Samples for the *in-gel* fluorescence of the constructs Mistic-TM13, Mistic-TM45, Mistic-TM56, Mistic-TM67 were prepared as described elsewhere.[31] In-gel fluorescence was detected using the Fujifilm LA300 imaging system. The same gel was then incubated in presence of a Ni-NTA Atto Complex (Sigma) binding the His tag of the GFP fusion, following instructions from the manufacturer. His-tagged proteins were visualized using the

Fujifilm LAS-3000 imaging system (λ_{em} 669, λ_{ex} white light).

5.7 Samples preparation and NMR spectroscopy: NMR samples (TM45, N-TM45, TM56 or TM67) were prepared by dissolving the lyophilized powder obtained after the reverse-phase C4 purification in the final NMR buffer (200 µl of 90% H_2O/D_2O containing 50 mM KPi, pH 6.0) in presence of selected detergent (5% wt/vol. of either LPPG, DPC, DHPC, LMPC or LMPG). The solution was subjected to several cycles of water-bath sonication and vigorous shaking at 37 °C until the powder was completely dissolved. The solution was then centrifuged at 10'000 g for ten minutes and the supernatant was transferred into a Shigemi NMR tube (Shigemi Inc.) for NMR measurements. The final protein concentrations were 0.2 mM and 0.4 mM for the ^{15}N and the [^{15}N,^{13}C,^{2}H] samples, respectively.

We recently discovered that the exact details of sample preparation have a substantial impact on the quality of spectra. Often better samples yielding superior spectra were obtained when using a procedure suggested by Killian *et al*.[66] Therein the lyophilized protein was dissolved in hexafluoroisopropanol (HFIP, 50 µL) resulting in a clear solution. To this, NMR sample buffer (200µl of 40 mM phosphate, 20 mM DTT, 10% D_2O, 1 mM TMSP and 5% LPPG) was added. The resulting cloudy suspension was diluted to 1 ml with water and the solution shaken at 37 °C with sporadic sonication cycles until a clear solution was obtained (4-8 h). The latter was lyophilized, redissolved in water and sonicated at 25 °C until a clear solution was obtained, which was lyophilized again. This step was repeated three times. Finally, the lyophilized powder was dissolved in water (200 µL) containing DTT (20 mM).

All spectra were recorded on Bruker AV-600 or AV-700 spectrometers equipped with cryoprobes at 320K. The 2D experiments utilized gradient-selected coherence selection (echo-antiecho) [67] in combination with sensitivity enhancement schemes. Proton-nitrogen correlation maps were measured as [^{15}N,^{1}H]-HSQC [68] (non-deuterated proteins) or as TROSY [69] (deuterated proteins) experiments. For assignment purposes of the receptor resonances, TROSY-versions [70] of standard 3D triple-resonance experiments

were recorded, such as the HNCO, HN(CA)CO, HNCA, HN(CO)CACB or the HNCACB [38,71] as well as from ^{15}N-resolved NOESY experiments. For more details on data acquisition see the Supp. Mat. Table S2.

Raw data was processed using the Bruker Topspin software version 2.1 and transferred to CARA[72] for further analysis. To investigate the effect of point mutations within TM4 on the signal intensity of signals from residues in TM4, a few well-resolved resonances were integrated. To eliminate influences of different concentrations or oligomeric states all integrals were normalized to obtain same integrals for G258, a flexible residue from the C-terminal tail.

6. Acknowledgements

We acknowledge financial support by the Swiss National Science Foundation (grant No 31003A_124649) and the OPO Foundation. The DNA sequence of Mistic was obtained from the group of Prof S. Choe (Salk, San Diego). We like to thank Dr. E. Geertsma for general support on this project and Prof. R. Dutzler for the possibility to use part of his lab equipment.

7. Table of Contents

To facilitate the assignment of the entire Y4 receptor or in order to enable study of its folding the almost complete sequence of the GPCR is expressed in large fragments comprising 2-3 transmembrane helices. Solution NMR techniques reveal that secondary structure in these fragments to a very large extend coincides with predictions for the entire receptor. Challenges and prospects for further studies of GPCR by NMR are discussed.

8. References:

[1] K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, M. Miyano, *Science* **2000**, *289*, 739-745.

[2] T. Warne, M. J. Serrano-Vega, J. G. Baker, R. Moukhametzianov, P. C. Edwards, R. Henderson, A. G. Leslie, C. G. Tate, G. F. Schertler, *Nature* **2008**, *454*, 486-491.

[3] S. G. Rasmussen, H. J. Choi, D. M. Rosenbaum, T. S. Kobilka, F. S. Thian, P. C. Edwards, M. Burghammer, V. R. Ratnala, R. Sanishvili, R. F. Fischetti, G. F. Schertler, W. I. Weis, B. K. Kobilka, *Nature* **2007**, *450*, 383-387.

[4] V. P. Jaakola, M. T. Griffith, M. A. Hanson, V. Cherezov, E. Y. Chien, J. R. Lane, A. P. Ijzerman, R. C. Stevens, *Science* **2008**, *322*, 1211-1217.

[5] E. Y. Chien, W. Liu, Q. Zhao, V. Katritch, G. W. Han, M. A. Hanson, L. Shi, A. H. Newman, J. A. Javitch, V. Cherezov, R. C. Stevens, *Science* **2010**, *330*, 1091-1095.

[6] B. Wu, E. Y. Chien, C. D. Mol, G. Fenalti, W. Liu, V. Katritch, R. Abagyan, A. Brooun, P. Wells, F. C. Bi, D. J. Hamel, P. Kuhn, T. M. Handel, V. Cherezov, R. C. Stevens, *Science* **2010**, *330*, 1066-1071.

[7] T. Shimamura, M. Shiroishi, S. Weyand, H. Tsujimoto, G. Winter, V. Katritch, R. Abagyan, V. Cherezov, W. Liu, G. W. Han, T. Kobayashi, R. C. Stevens, S. Iwata, *Nature* **2011**, *475*, 65-70.

[8] G. Lebon, T. Warne, P. C. Edwards, K. Bennett, C. J. Langmead, A. G. Leslie, C. G. Tate, *Nature* **2011**, *474*, 521-525.

[9] D. M. Rosenbaum, C. Zhang, J. A. Lyons, R. Holl, D. Aragao, D. H. Arlow, S. G. Rasmussen, H. J. Choi, B. T. Devree, R. K. Sunahara, P. S. Chae, S. H. Gellman, R. O. Dror, D. E. Shaw, W. I. Weis, M. Caffrey, P. Gmeiner, B. K. Kobilka, *Nature* **2011**, *469*, 236-240.

[10] T. Warne, R. Moukhametzianov, J. G. Baker, R. Nehme, P. C. Edwards, A. G. Leslie, G. F. Schertler, C. G. Tate, *Nature* **2011**, *469*, 241-244.

[11] F. Xu, H. Wu, V. Katritch, G. W. Han, K. A. Jacobson, Z. G. Gao, V. Cherezov, R. C. Stevens, *Science* **2011**, *332*, 322-327.

[12] J. Standfuss, P. C. Edwards, A. D'Antona, M. Fransen, G. Xie, D. D. Oprian, G. F. Schertler, *Nature* **2011**, *471*, 656-660.

[13] S. G. Rasmussen, B. T. DeVree, Y. Zou, A. C. Kruse, K. Y. Chung, T. S. Kobilka, F. S. Thian, P. S. Chae, E. Pardon, D. Calinski, J. M. Mathiesen, S. T. Shah, J. A. Lyons, M. Caffrey, S. H. Gellman, J. Steyaert, G. Skiniotis, W. I. Weis, R. K. Sunahara, B. K. Kobilka, *Nature* **2011**, *477*, 549-555.

[14] J. H. Park, P. Scheerer, K. P. Hofmann, H. W. Choe, O. P. Ernst, *Nature* **2008**, *454*, 183-187.

[15] P. Scheerer, J. H. Park, P. W. Hildebrand, Y. J. Kim, N. Krauss, H. W. Choe, K. P. Hofmann, O. P. Ernst, *Nature* **2008**, *455*, 497-502.

[16] R. M. Bill, P. J. Henderson, S. Iwata, E. R. Kunji, H. Michel, R. Neutze, S. Newstead, B. Poolman, C. G. Tate, H. Vogel, *Nat. Biotechnol.* **2011**, *29*, 335-340.

[17] A. Gautier, J. P. Kirkpatrick, D. Nietlispach, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 7297-7300.

[18] A. Gautier, H. R. Mott, M. J. Bostock, J. P. Kirkpatrick, D. Nietlispach, *Nat. Struct. Mol. Biol.* **2010**, *17*, 768-774.

[19] S. Reckel, D. Gottstein, J. Stehle, F. Löhr, M. K. Verhoefen, M. Takeda, R. Silvers, M. Kainosho, C. Glaubitz, J. Wachtveitl, F. Bernhard, H. Schwalbe, P. Güntert, V. Dötsch, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 11942-11946.

[20] W. D. Van Horn, H. J. Kim, C. D. Ellis, A. Hadziselimovic, E. S. Sulistijo, M. D. Karra, C. Tian, F. D. Sonnichsen, C. R. Sanders, *Science* **2009**, *324*, 1726-1729.

[21] A. Arora, F. Abildgaard, J. H. Bushweller, L. K. Tamm, *Nat. Struct. Biol.* **2001**, *8*, 334-338.

[22] P. M. Hwang, W. Y. Choy, E. I. Lo, L. Chen, J. D. Forman-Kay, C. R. Raetz, G. G. Prive, R. E. Bishop, L. E. Kay, *Proc. Natl. Acad. Sci. U S A* **2002**, *99*, 13560-13565.

[23] A. Neumoin, L. S. Cohen, B. Arshava, S. Tantry, J. M. Becker, O. Zerbe, F. Naider, *Biophys. J.* **2009**, *96*, 3187-3196.

[24] C. Zou, S. Kumaran, S. Markovic, R. Walser, O. Zerbe, *ChemBioChem* **2008**, *9*, 2276-2284.

[25] C. Zou, F. Naider, O. Zerbe, *J. Biomol. NMR* **2008**, *42*, 257-269.

[26] T. P. Roosild, J. Greenwald, M. Vega, S. Castronovo, R. Riek, S. Choe, *Science* **2005**, *307*, 1317-1321.

[27] D. M. LeMaster, *Methods Enzymol.* **1989**, *177*, 23-43.

[28] W. D. Morgan, A. Kragt, J. Feeney, *J. Biomol. NMR* **2000**, *17*, 337-347.

[29] T. Etezady-Esfarjani, S. Hiller, C. Villalba, K. Wüthrich, *J. Biomol. NMR* **2007**, *39*, 229-238.

[30] D. Drew, D. Sjostrand, J. Nilsson, T. Urbig, C. N. Chin, J. W. de Gier, G. von Heijne, *Proc. Natl. Acad. Sci. U S A* **2002**, *99*, 2690-2695.

[31] E. R. Geertsma, M. Groeneveld, D. J. Slotboom, B. Poolman, *Proc. Natl. Acad. Sci. U S A* **2008**, *105*, 5722-5727.

[32] D. E. Warschawski, A. A. Arnold, M. Beaugrand, A. Gravel, E. Chartrand, I. Marcotte, *Biochim. Biophys. Acta* **2011**, *1808*, 1957-1974.

[33] T. K. Ritchie, Y. V. Grinkova, T. H. Bayburt, I. G. Denisov, J. K. Zolnerciks, W. M. Atkins, S. G. Sligar, *Methods Enzymol.* **2009**, *464*, 211-231.

[34] G. G. Prive, *Methods* **2007**, *41*, 388-397.

[35] R. C. Page, J. D. Moore, H. B. Nguyen, M. Sharma, R. Chase, F. P. Gao, C. K. Mobley, C. R. Sanders, L. Ma, F. D. Sönnichsen, S. Lee, S. C. Howell, S. J. Opella, T. A. Cross, *J. Struct. Funct. Gen.* **2006**, *7*, 51-64.

[36] R. D. Krüger-Koplin, P. L. Sorgen, S. T. Krüger-Koplin, I. O. Rivera-Torres, S. M. Cahill, D. B. Hicks, L. Grinius, T. A. Krulwich, M. E. Girvin, *J. Biomol. NMR* **2004**, *28*, 43-57.

[37] X. Shan, K. H. Gardner, D. R. Muhandiram, N. S. Rao, C. H. Arrowsmith, L. E. Kay, *J. Am. Chem. Soc.* **1996**, *118*, 6570-6579.

[38] M. Sattler, J. Schleucher, C. Griesinger, *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93-158.

[39] L. S. Cohen, B. Arshava, A. Neumoin, J. M. Becker, P. Güntert, O. Zerbe, F. Naider, *Biochim. Biophys. Acta* **2011**, *1808*, 2674-2684.

[40] A. Senes, D. C. Chadi, P. B. Law, R. F. Walters, V. Nanda, W. F. Degrado, *J. Mol. Biol.* **2007**, *366*, 436-448.

[41] W. C. Wimley, S. H. White, *Nature Struct. Biol.* **1996**, *3*, 842-848.

[42] A. N. Ridder, S. Morein, J. G. Stam, A. Kuhn, B. de Kruijff, J. A. Killian, *Biochemistry* **2000**, *39*, 6521-6528.

[43] H. J. Kim, S. C. Howell, W. D. Van Horn, Y. H. Jeon, C. R. Sanders, *Prog. Nucl. Magn. Reson. Spectrosc.* **2009**, *55*, 335-360.

[44] S. P. Zhang, G. Zubay, E. Goldman, *Gene* **1991**, *105*, 61-72.

[45] D. Drew, M. Lerch, E. Kunji, D. J. Slotboom, J. W. de Gier, *Nat. Methods* **2006**, *3*, 303-313.

[46] P. Schmidt, C. Berger, H. A. Scheidt, S. Berndt, A. Bunge, A. G. Beck-Sickinger, D. Huster, *Biophys. Chem.* **2010**, *150*, 29-36.

[47] A. M. Seddon, P. Curnow, P. J. Booth, *Biochim. Biophys. Acta* **2004**, *1666*, 105-117.

[48] K. Lundström, *Trends Biotechnol.* **2005**, *23*, 103-108.

[49] K. Michalke, M. E. Graviere, C. Huyghe, R. Vincentelli, R. Wagner, F. Pattus, K. Schroeder, J. Oschmann, R. Rudolph, C. Cambillau, A. Desmyter, *Anal. Biochem.* **2009**, *386*, 147-155.

[50] K. Michalke, C. Huyghe, J. Lichiere, M. E. Graviere, M. Siponen, G. Sciara, I. Lepaul, R. Wagner, C. Magg, R. Rudolph, C. Cambillau, A. Desmyter, *Anal. Biochem.* **2010**, *401*, 74-80.

[51] J. L. Baneres, J. L. Popot, B. Mouillac, *Trends Biotechnol.* **2011**, *29*, 314-322.

[52] V. Sarramegna, I. Muller, A. Milon, F. Talmont, *Cell Mol. Life Sci.* **2006**, *63*, 1149-1164.

[53] K. E. Caroccia, R. Estephan, L. S. Cohen, B. Arshava, M. Hauser, O. Zerbe, J. M. Becker, F. Naider, *Biopolymers* **2011**, *1808*, 2674-2684.

[54] L. S. Cohen, B. Arshava, R. Estephan, J. Englander, H. Kim, M. Hauser, O. Zerbe, M. Ceruso, J. M. Becker, F. Naider, *Biopolymers* **2008**, *90*, 117-130.

[55] R. Estephan, J. Englander, B. Arshava, K. L. Samples, J. M. Becker, F. Naider, *Biochemistry* **2005**, *44*, 11795-11810.

[56] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J. Biomol. NMR* **2009**, *44*, 213-223.

[57] D. S. Wishart, B. D. Sykes, *Methods Enzymol.* **1994**, *239*, 363-392.

[58] J. Marino, E. R. Geertsma, O. Zerbe, *manuscript in preparation*

[59] A. Neumoin, B. Arshava, J. Becker, O. Zerbe, F. Naider, *Biophys. J.* **2007**, *93*, 467-482.

[60] C. G. Tate, G. F. Schertler, *Curr. Opin. Struct. Biol.* **2009**, *19*, 386-395.

[61] D. M. Engelman, Y. Chen, C. N. Chin, A. R. Curran, A. M. Dixon, A. D. Dupuy, A. S. Lee, U. Lehnert, E. E. Matthews, Y. K. Reshetnyak, A. Senes, J. L. Popot, *FEBS Lett.* **2003**, *555*, 122-125.

[62] J. L. Popot, D. M. Engelman, *Biochemistry* **1990**, *29*, 4031-4037.

[63] Sambrook, J., Russel, D. W. *Molecular Cloning: a laboratory manual*; Cold Spring Harbor Press: New York, 1989;

[64] E. R. Geertsma, R. Dutzler, *Biochemistry* **2011**, *50*, 3272-3278.

[65] B. Miroux, J. E. Walker, *J. Mol. Biol.* **1996**, *260*, 289-298.

[66] J. A. Killian, T. P. Trouard, D. V. Greathouse, V. Chupin, G. Lindblom, *FEBS Lett.* **1994**, *348*, 161-165.

[67] L. E. Kay, P. Keifer, T. Saarinen, *J. Am. Chem. Soc,* **1992**, *114*, 10663-10665.

[68] G. Bodenhausen, D. J. Ruben, *Chem. Phys. Lett.* **1980**, *69*, 185-189.

[69] K. Pervushin, R. Riek, G. Wider, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12366-12371.

[70] M. Salzmann, G. Wider, K. Pervushin, H. Senn, K. Wüthrich, *J. Am. Chem. Soc.* **1999**, *121*, 844.

[71] T. Yamazaki, W. Lee, C. H. Arrowsmith, D. R. Muhandiram, L. E. Kay, *J. Am. Chem. Soc.* **1994**, *116*, 11655-11666.

[72] R. Keller, **2004**, Catina Verlag Goldau.

9. Supplementary Material

Table S1: Oligomers used for cloning of various receptor fragments

Name	Sequence (5-3')*
TM13CO-NdeI/His10-For	aat tta cat atg cat cac cat cac cat cac cat cac atgaataccagccatctgctg
TM13CO-BamHi-Rev	aat taa gga tcc *tta* ctggctaatgctcggtttc
TM13WT- NdeI-His10-For	aat tta cat atg cat cac cat cac cat cac cat cac atgaacacctctcacctcct
TM13WT-BamHi-Rev	aat taa gga tcc *tta* ctgtgagatgctgggcttc
TM45CO- NdeI/His10-For	aat tta cat atg cat cac cat cac cat cac cat cac gaacgtcatcagctgattat
TM45CO- BamHI -Rev	aat taa gga tcc *tta* cagaacaacattaacctgtt
TM45 WT- NdeI/His10For	aat tta cat atg cat cac cat cac cat cac cat cac gagaggcatcagctcatcatc
TM45WT- BamHI -Rev	aat taa gga tcc *tta* cagcaccacattgacctg
TM56CO-NdeI/His10-For	aat tta cat atg cat cac cat cac cat cac cat cac attctggaaaatgtgtttca
TM56CO- BamHI -Rev	aat taa gga tcc *tta* attgccatgacaaatcggaa
TM56WT- NdeI/His10-For	aat tta cat atg cat cac cat cac cat cac cat cac atcctggagaatgtcttc
TM56WT- BamHI -Rev	aat taa gga tcc *tta* gttcccgtggcagatgggga
TM67-3C-For	gaa gcg gaa aaa ctg ttt aac cag ctg gaa gtg ctg ttc cag ggg ccc
TM67-20aa-linker-For	cgc gcg aag cag ttc gtg gaa gat aac gcg gaa aac gat gaa gcg gaa aaa ctg ttt aac
TM67-HindIII-His10-for	gac aag ctt cac cat cac cat cac cat cac cat cac cat gaa cgc cgcgcgaagcagttcgtggaagat
TM67-For	ctggaagtgctgttccaggggcccggcgcctgcagaggcaggg
TM67-Rev	aat taa ggatcc *tta* aatgggattggacctgcca
Mistic-For	tttttaa catatg ttttgtacatttttgaaa
Mistic-Rev	ttaa ggatcc ttctttttcgccttcttc
TM13-wt-3C-For	aattggatccctggaagttctgtttcagggtccgatgaacacctctcacctcc
Mistic-FX-For	atatatgctcttctagtttctgtacatttttgaaaaacatcac
TM13-FX-Rev	tatatagctcttcatgcctgtgagatactgggtttccagcctgt
TM45-FX-Rev	tatatagctcttcatgcacccttgtgaaacacgcgcccctgcct
TM56-FX-Rev	tatatagctcttcatgcgcagatggggatggcctcatggtgc
TM67-FX-Rev	tatatagctcttcatgcaatgggattggacctgccacttagcct

*Restrictions sites used for cloning of PCR products are depicted in red; octa-His tag introduced *via* forward primers is depicted in blue; 3C protease site used for constructs ΔTrpLE-TM67 and Mistic-TM13 is depicted in purple; the 20 aa linker used for ΔTrpLE-TM67 is depicted in green.

Figure S1: Solubilization assay: 15% acrylamide SDS-page, blotted on a PVC membrane and detected via anti-His antibody. Lysed cells were incubated with 1% (wt/vol) of LDAO (1), DPC (2) or LPPG (3). Lanes 4-6 correspond to TM13 solubilized by LDAO (4), DPC (5) or LPPG (6).

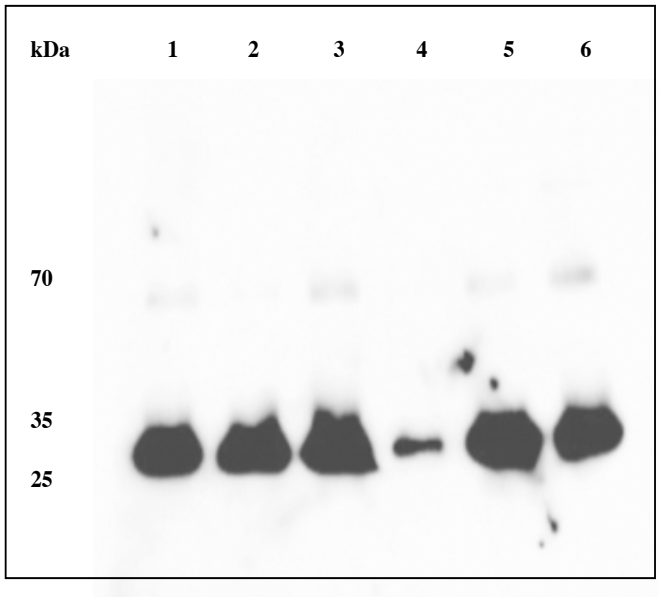


Figure S2: Locations of segments of the Y4 GPCR investigated in this work. Residue numbering is indicated a certain positions, and the first and last TM helix residues are encircled in bold black. Both full-sequence numbering as well a numbering according to the Ballesteros-Weinstein system (J. Ballesteros, H. Weinstein, *Methods Neurosci.* **1995**, *25*, 366-428) (in brackets) is used. The location of position 50 is highlighted by bold red circles.

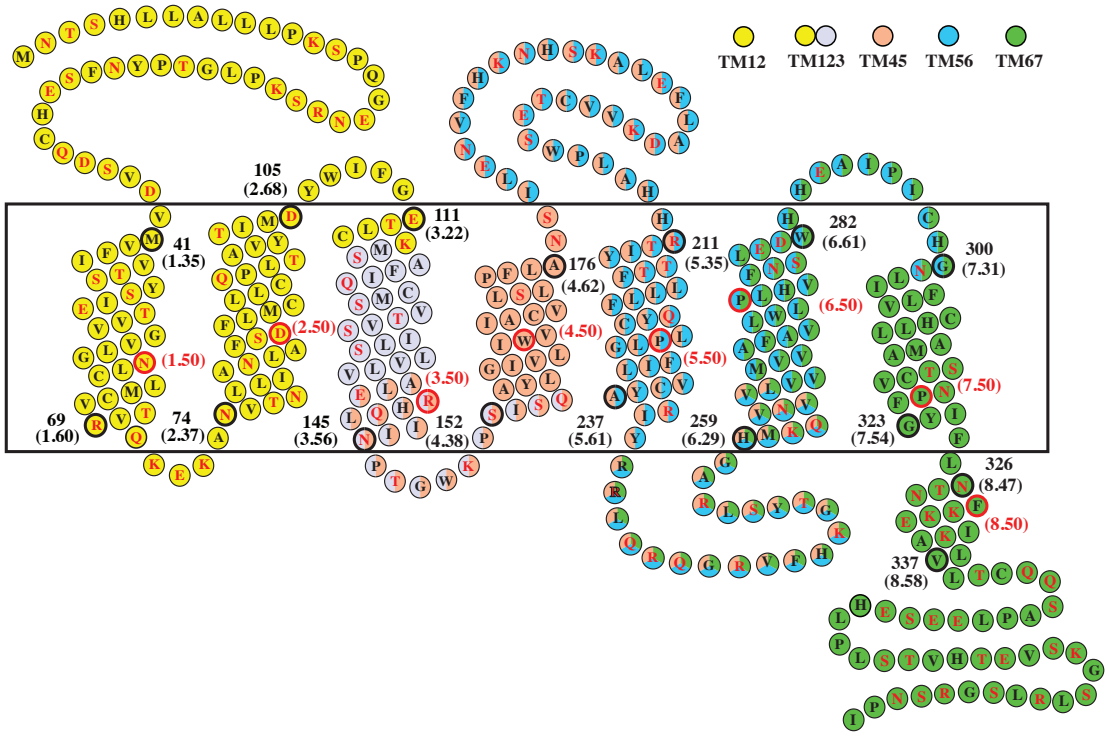


Table S2: Spectroscopic details of NMR experiments:

TM45								
experiment	# of data points			spectral width			scans	Remarks
	F3	F2	F1	F3	F2	F1		
[^{15}N,^{1}H]-HSQC		2048	256		16	32	128	
HNCO	2048	40	128	14	27	22	16	
HN(CO)CACB	2048	40	128	14	30	70	16	
HNCA	2048	40	128	16	30	32	16	
HN(CO)CA	2048	40	128	16	30	32	16	
^{15}N-NOESY	2048	40	200	14	27	10	16	100ms mix
NTM45								
	Data points			Spectral width			NS	
	F3	F2	F1	F3	F2	F1		
[^{15}N,^{1}H]-HSQC		2048	300		16	33	32	
HNCO	2048	44	140	14	33	22	8	
HN(CA)CO	2048	48	128	16	25	16	8	
HNCACB	2048	44	140	18	25	70	16	
HN(CO)CACB	2048	44	140	18	25	70	16	
HNCA	2048	128	40	18	32	24	16	
HN(CO)CA	2048	40	128	18	24	32	16	
^{15}N-NOESY	2048	42	180	14	24	10	16	150ms mix
TM45TFE								
	Data points			Spectral width			NS	
	F3	F2	F1	F3	F2	F1		
[^{15}N,^{1}H]-HSQC		2048	256		16	36	8	
HNCO	2048	40	128	14	24	22	4	
HNCACO	2048	40	128	14	24	22	16	
HNCACB	2048	40	140	14	24	70	8	
HN(CO)CBCA	2048	40	128	14	24	70	8	
^{15}N-NOESY	2048	50	200	14	30	10	16	120ms mix
TM67								
	Data points			Spectral width			NS	
	F3	F2	F1	F3	F2	F1		
[^{15}N,^{1}H]-HSQC		2048	256		18	33	32	
HNCO	2048	40	128	18	22	22	8	
HNCACO	2048	40	110	18	22	22	16	
HNCACB	2048	40	124	18	22	70	32	
HN(CO)CACB	2048	40	128	18	22	70	16	
HNCA	2048	40	92	18	22	20	32	
HN(CO)CA	2048	40	97	18	22	18	32	
^{15}N-NOESY	2048	40	161	14	22	10	16	120ms
TM12								
Expt	Data points			Spectral width			NS	
	F3	F2	F1	F3	F2	F1		
[^{15}N,^{1}H]-HSQC		2048	256		18	25.7	16	
HNCO	2048	40	128	18	26	22	8	
HNCACO	2048	40	128	18	25	22.6	16	
HNCACB	2048	40	150	18	26	60	32	
HN(CO)CACB	2048	40	128	16	25	65	32	
HNCA	2048	40	128	16	25	32	8	
HN(CO)CA	2048	40	128	16	25	32	16	
^{15}N-NOESY	2048	40	200	14	24	10	16	100ms

Table S3: Assignment statistics of [$^{15}N^{13}C^{2}H$]-NTM45 in LPPG.

NTM45	Amino acids number	Assignment in sequence	Assign	Assign %
N-term	50 (8 Histag, 4 Proline)	MHHHHHHHHH**MNTSHLLALLL**P**KS**P**QGENRSK**P**LGT**PYNFSEHCQDSVDVM	23	60%
Half-I2 loop	9 (1 Proline)	**GWK**PSISQA	3	37%
TM4	22 (1 Proline)	Y**LG**IVLIWVIACVLSLPFLANS	2	1%
E2 loop	38 (1 Proline)	ILENVFHKNHSKAL**EFLADKVVCTESW**PLAHHRTIY**TT**	15	40%
TM5	22 (1 Proline)	**FLLLFQYCL**P**LGFILVCYARIY**	21	100%
Half-I3 loop	24	**RRLQRQGRVFHKGTYSLRAGHMKQ**	24	100%
Total	165 (8 Histag, 8 Proline)	MHHHHHHHHH**MNTSHLLALLL**P**KS**P**QGENRSK**P**LGT**PYNFSEHCQDSVDVM**GWK**PSISQAY**LG**IVLIWVIACVLSLPFLANSILENVFHKNHSKAL**EFLADKVVCTESW**PLAHHRTIY**TTFLLLFQYCL**P**LGFILVCYARIYRRLQRQGRVFHKGTYSLRAGHMKQ**	88	59%

Table S4 Assignment statistics of [$^{15}N^{13}C$]-TM45 in LPPG.

TM45	Amino acids number	Assignment in sequence	Assign	Assign %
I2 loop	28 (8 Histag, 2 Proline)	MHHHHHHHHH**ERHQLIIN**P**TGWK**PSISQA	12	67%
TM4	22 (1 Proline)	Y**LGIV**LIWVIACVLSLP**FLANS**	8	38%
E2 loop	38 (1 Proline)	ILENVFHKNHSKAL**EFLADKVVCTESW**PLAHHRTIY**TT**	25	67%
TM5	22 (1 Proline)	**FLLLFQYCL**P**LGFILVCYARIY**	21	100%
I3 loop	29	**RRLQRQGRVFHKGTYSLRAGHMKQVNVVL**	29	100%
Total	139 (8 Histag, 5 Proline)	MHHHHHHHHH**ERHQLIIN**P**TGWK**PSISQAY**LGIV**LIWVIACVLSLP**FLANS**ILENVFHKNHSKAL**EFLADKVVCTESW**PLAHHRTIY**TTFLLLFQYCL**P**LGFILVCYARIYRRLQRQGRVFHKGTYSLRAGHMKQVNVVL**	95	75%

Table S5 Assignment statistics of [$^{15}N^{13}C$]-NTM45 in TFE:H_2O.

NTM45	Amino acids number	Assignment in sequence	Assign	Assign %
N-term	50 (4 Proline)	M**HHHHHHHHHMNTSHLLALL LPKSPQGENRSK**P**LGT**P**YNFS EHCQDSVDVM**	45	100%
Half-I2 loop	9 (1 Proline)	**GWK**P**SISQA**	8	100%
TM4	22 (1 Proline)	**YLGIVLIWVIACVLSL**P**FLANS**	21	100%
E2 loop	38 (1 Proline)	**I**LE**NVFHKN**HSK**AL**EFL**AD**KV V**CTESW**P**LAHHRTIYTT**	26	70%
TM5	22 (1 Proline)	**F**LLL**FQYCL**P**LGFILVCYARIY**	18	85%
Half-I3 loop	24	**RRLQRQGRVFHKGTYSLRAG HMKQ**	24	100%
Total	165 (8 Proline)	M**HHHHHHHHHMNTSHLLALL LPKSPQGENRSK**P**LGT**P**YNFS EHCQDSVDVMGWK**P**SISQAY LGIVLIWVIACVLSL**P**FLANSI** LE**NVFHKN**HSK**AL**EFL**AD**KVV **CTESW**P**LAHHRTIYTTF**LLL**F QYCL**P**LGFILVCYARIYRRLQ RQGRVFHKGTYSLRAGHMK Q**	142	90%

Table S6 Assignment statistics of [$^{15}N^{13}C$]-TM67 in LPPG.

TM67	Amino acids number	Assignment in sequence	Assign	Assign %
I3 loop	26 (1 Proline, 1st Gly not observed)	GP**RRLQRQGRVFHKGTYSL RAGHM**KQ	22	92%
TM6	22 (1 Proline)	**VNVV**L**VVM**V**VA**F**AV**L**W**LPL HVF	12	57%
E3 loop	17 (1 Proline)	**NSLEDWHHEAI**P**ICHGN**	17	100%
TM7	24 (1 Proline)	**LIFLVCHLLAMASTC** VNPF**IYGFL**	20	87%
C-terminus	50 (3 Proline)	**NTNFKKEIKALVLTCQQSA**P **LEESEHL**P**LSTVHTEVSKGS LRLSGRSN**P**I**	47	100%
Total	139 (7 Proline)	GP**RRLQRQGRVFHKGTYSL RAGHM**KQ**VNVV**L**VVM**V**VA** F**AV**L**W**LPLHVF**NSLEDWHH EAI**P**ICHGNLIFLVCHLLAM ASTC**VNPF**IYGFLNTNFKKEI KALVLTCQQSA**P**LEESEHL**P **LSTVHTEVSKGSLRLSGRSN** P**I**	118	90%

Figure S3: Detergent screen of TM67 using [^{15}N,^{1}H]-HSQC spectra and 300 μM ^{15}N-labeled TM67. Spectra are shown for samples in 5% DHPC (top left), 5% DMPG (top right), LMPC (bottom left) and LPPG (bottom right).

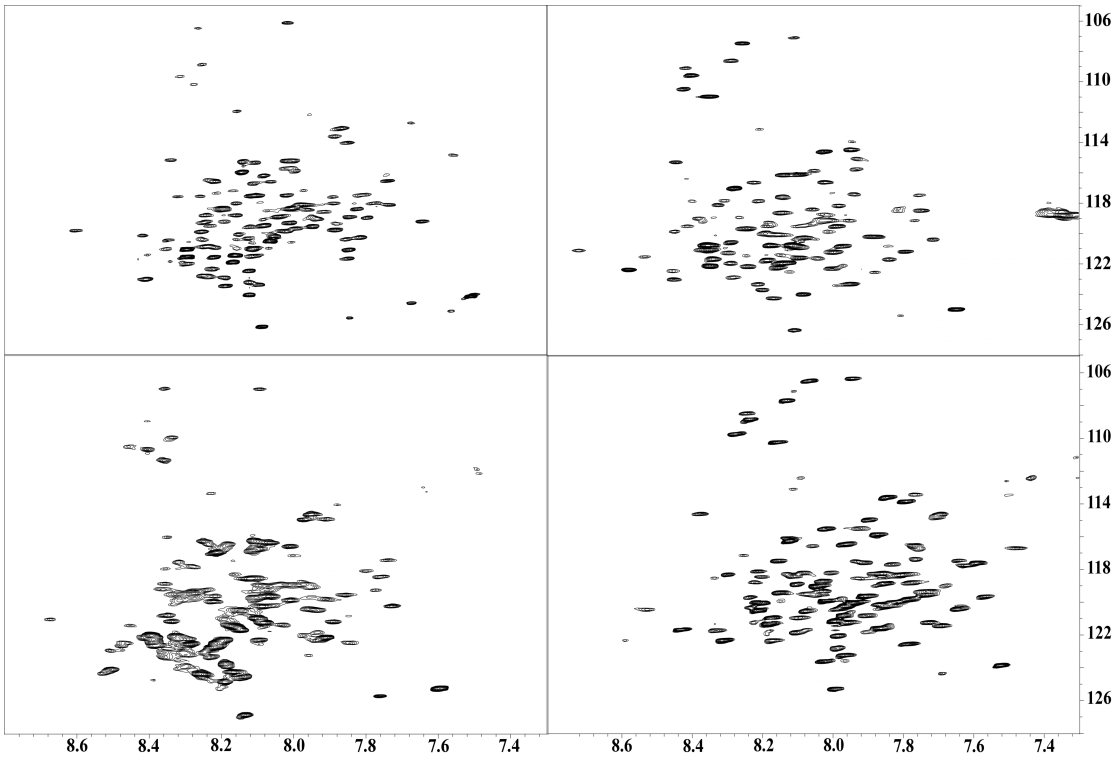


Figure S4: [^{15}N,^{1}H]-TROSY spectrum of NY4-TM1-TM2

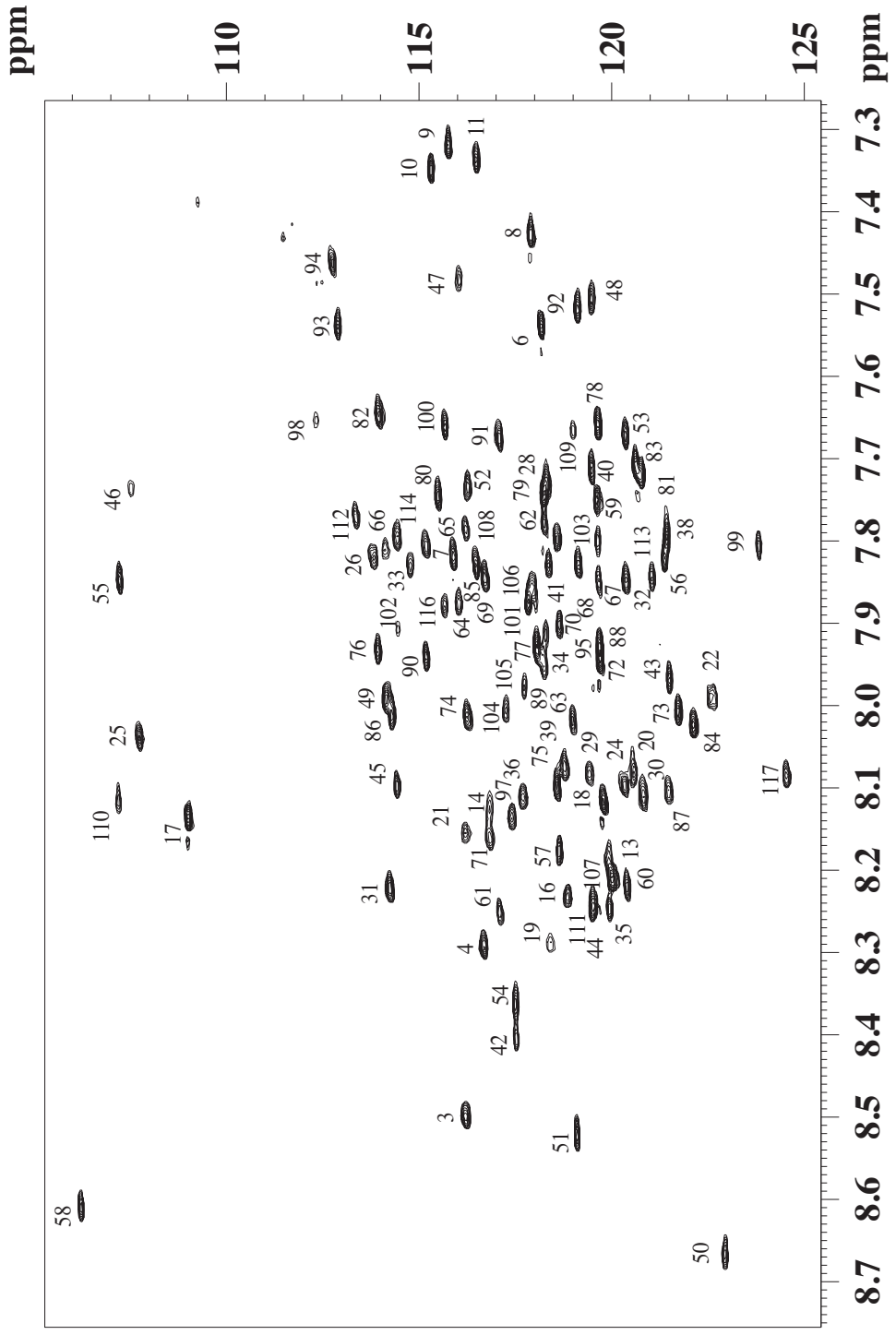


Figure S5: [^{15}N,^{1}H]-HSQC spectrum of TM4-TM5

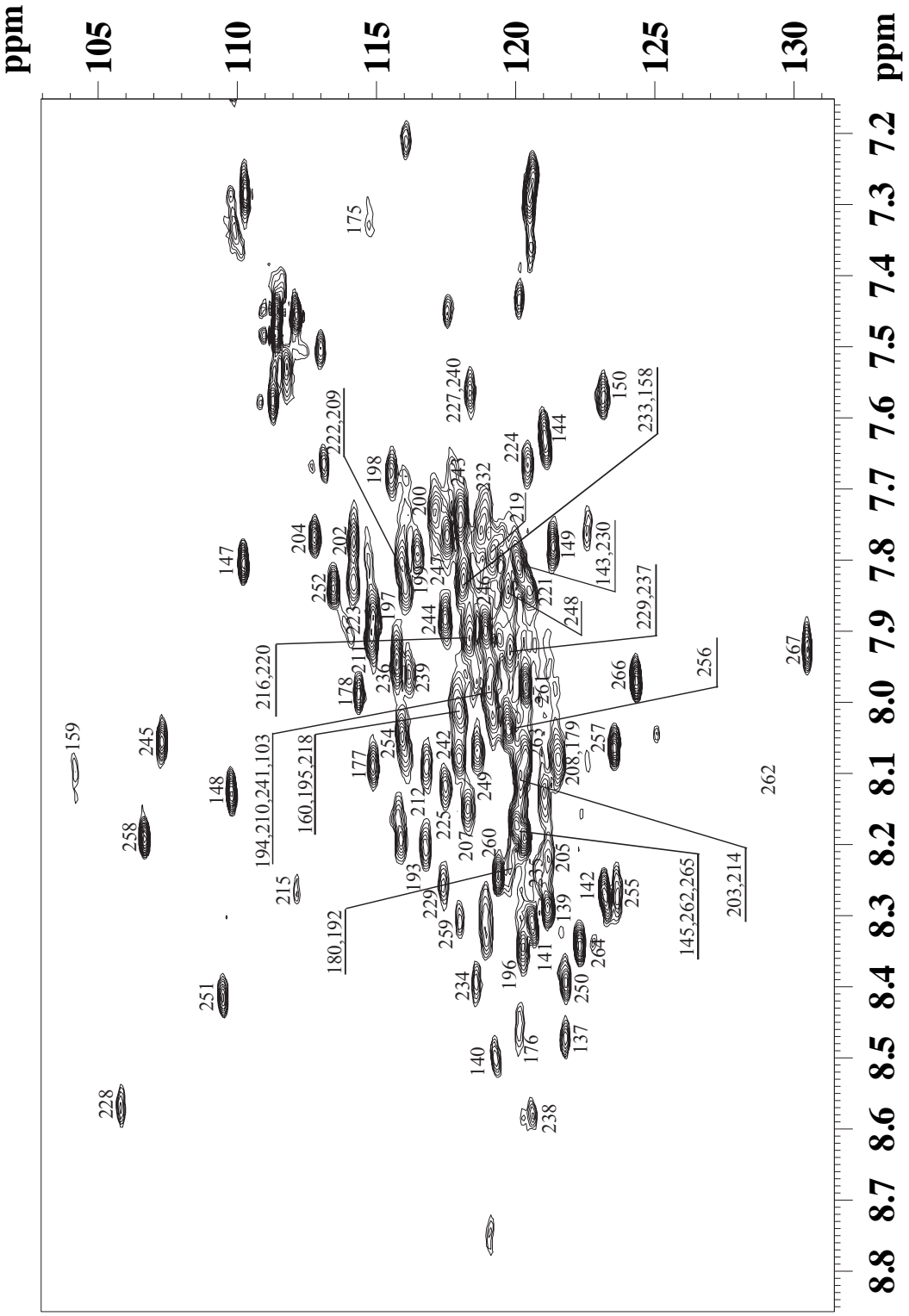


Figure S6: [^{15}N,^{1}H]-TROSY spectrum of TM6-TM7

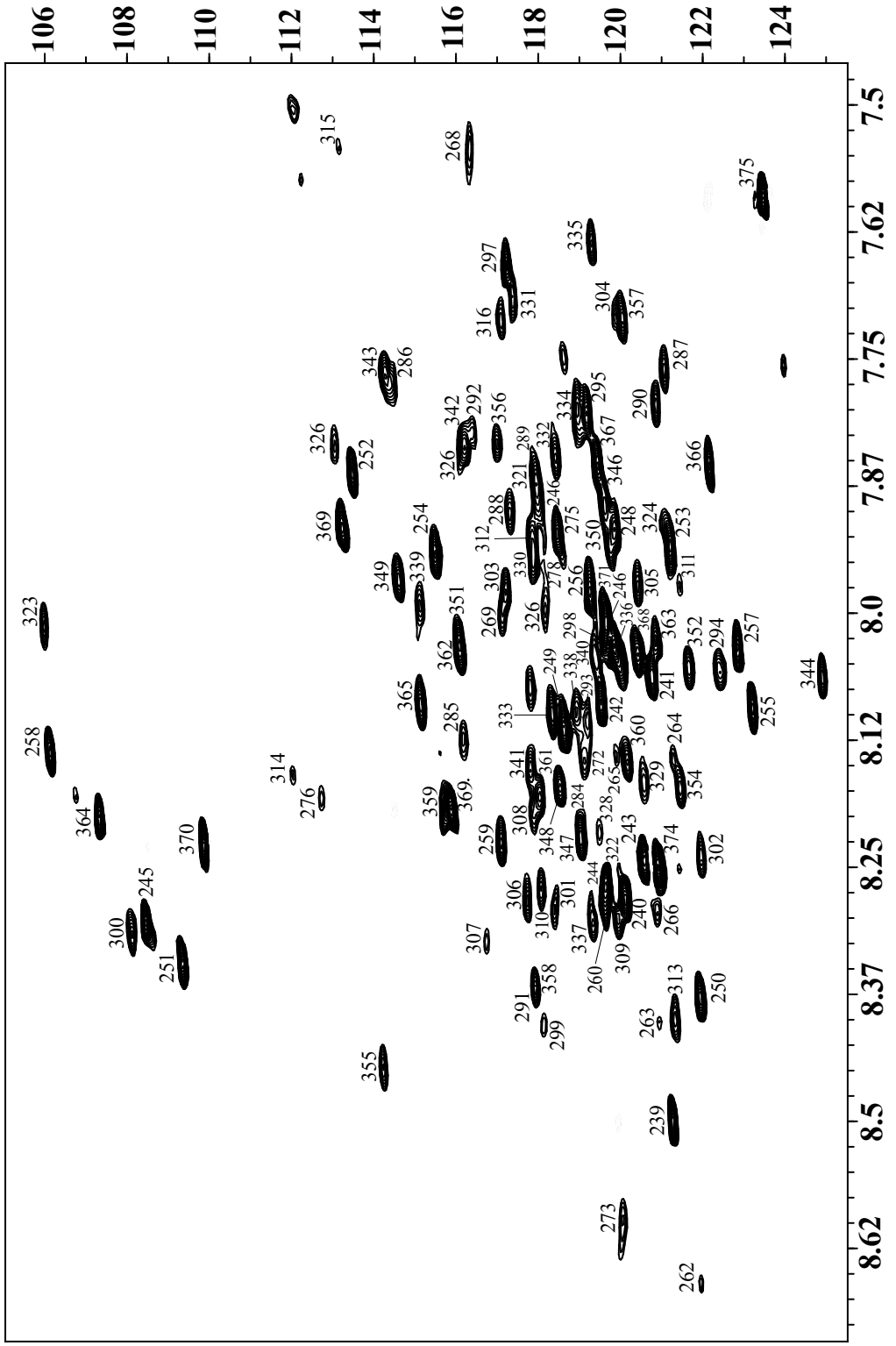


Figure S7: Prediction of secondary structure of TM45 in TFE/water based on backbone chemical shifts using the program TALOS+. Residues predicted to be helical are colored blue, those without secondary structure in orange.

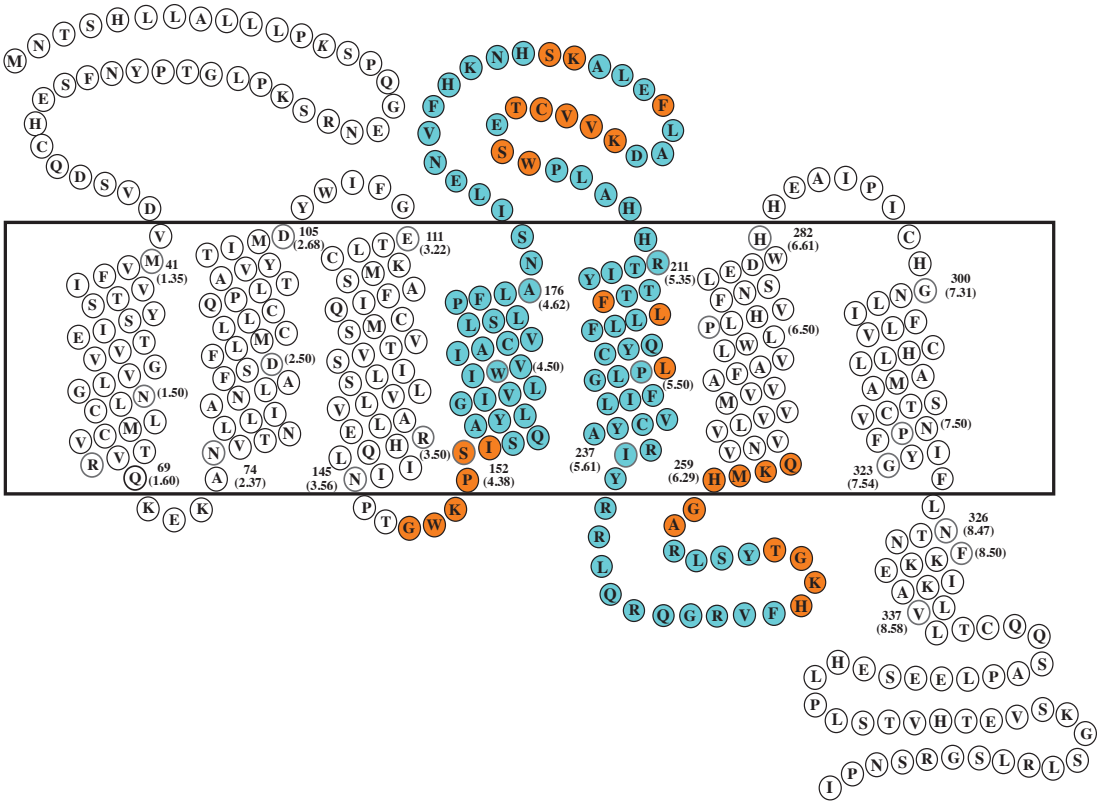
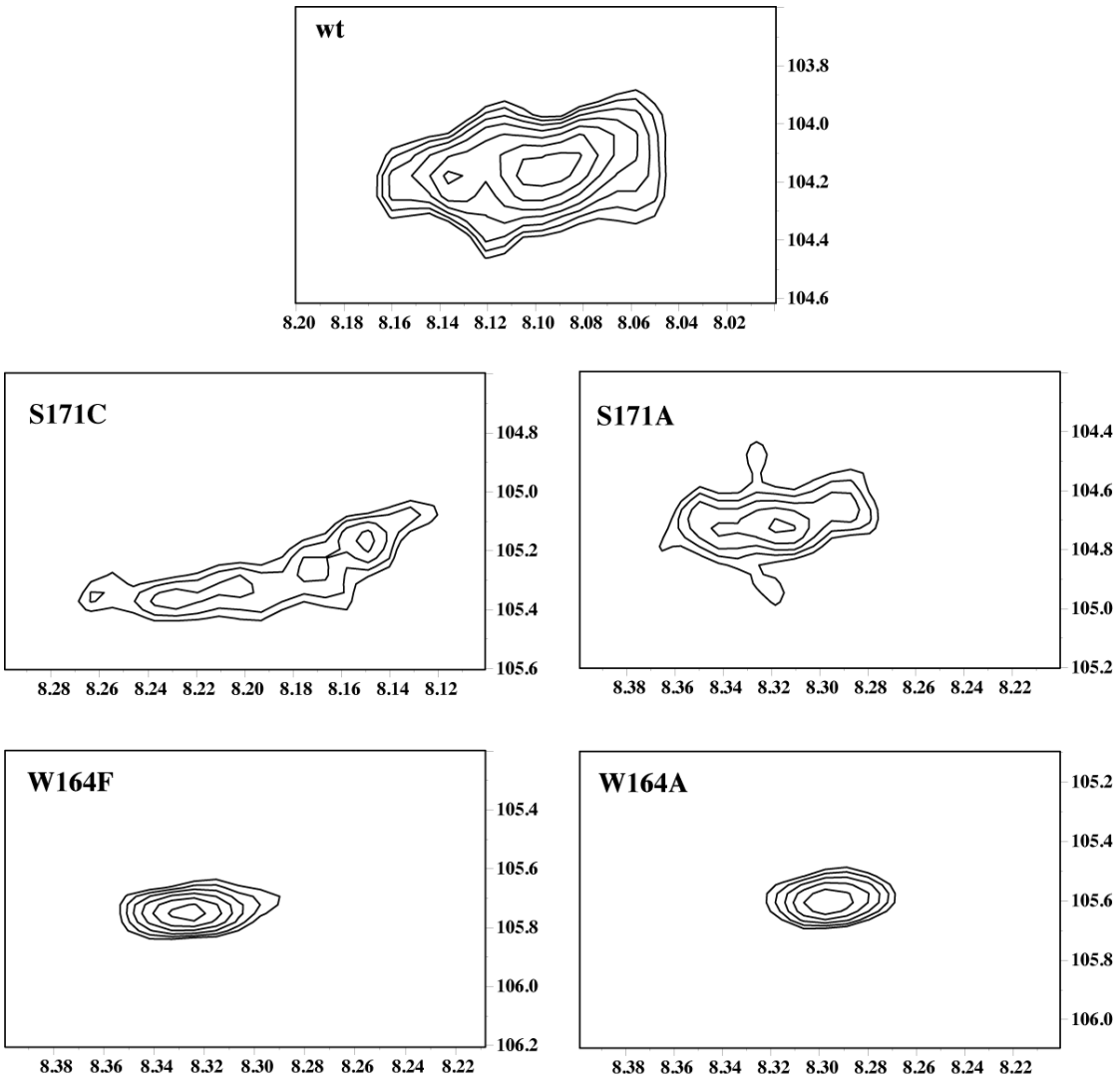


Figure S8: Expansions of [^{15}N,^{1}H]-HSQC spectra from TM45 (top), S171C and S171A (middle) and W164F or W164A (bottom) displaying the cross peak for G159 part of TM4.



Chapter III: Expression of C-terminal NPY4R fragments in inclusion bodies (IBs) and their purification

1. Abstract

Nowadays, transmembrane proteins (TMPs) are generally overexpressed by utilizing solubility-enhanced fusion tags such as SUMO, MBP, TRX, etc., which are easily cleaved by site-specific proteases and then followed by a simple affinity chromatographic purification. Alternatively, there are many other fusion protein tags existing for insoluble protein expression such as TrpΔLE, PurF, ketosteroid isomerase (KSI), and PagP, which require different conditions for cleavage like chemical cleavage, metal ion-catalyzed peptide bond cleavage etc., in presence of harsh denaturing conditions, followed by purification steps that require refolding. Expressing TMPs *directly* is widely practiced heterologously in non-native host systems such as *E.coli*, which often directs the TMP expression product into inclusion bodies due to the absence of ideal refolding machinery (such as chaperones), an overloading of the translocon machinery, and also because of the absence of a natural membrane environment that is capable of integrating multi-spanning hydrophobic domains. This last approach is many-a-times disadvantageous because of the total lack of overexpression, and thus the reason for choosing a fusion tag. As an advantage, it has to be noted that IBs are relatively homogenous and mostly comprise only negligible impurities[1].

In this chapter, we describe the successful expression and purification of a number of C-terminal fragments of NPY4R, in the form of inclusion bodies in *E.coli* either with or without the utilization of the fusion partner. Basically the fragments are 2-3 transmembrane in length with complete loops or termini in most cases. We describe one specific expression and purification protocol, which was suitable for purifying TMPs expressed as inclusion bodies in *E.coli* strain.

2. Introduction

Expression of heterologous proteins in *Escherichia coli* frequently results in forming masses of insoluble aggregates known as inclusion bodies. Although mechanisms behind

formation of inclusion bodies is not well understood, there are several reasons that could explain this phenomenon[1]: Solubility limitations, improper disulfide bond formation, the need to escape from protease sensitive enzymes, recombinant protein production at non-optimal pH and temperature especially during unfavorable conditions such as drought, an N-terminal aminoacid sequence that hampers rapid membrane integration, expression of protein at rates too fast for subsequent folding and membrane integration, and the lack of essential chaperones. Recent findings suggest that the formation of IBs can be compared to that of an amyloid fibril polymerization. This specificity was demonstrated by *in vitro* studies with the triggering of nucleation cores between protein monomers and then as microaggregates to form bigger IBs[2-5].

The biochemical and biophysical properties of IBs in *E. coli* have been comprehensively studied in the last 3 decades. They exist as porous ovoids or cylinders with diameter about 1 μm[6,7]. IBs of recombinantly expressed proteins are considered relatively pure. Proteomic investigations revealed that the IB aggregates are relatively homogenous to >95%[8-13]. Components that have been seldom detected bound to IBs are phospholipids, nucleic acids [11] and a small background of host cellular proteins that are co-isolated from inadequate cell disruption. Cytoplasmic proteins rarely observed bound to IBs are small heat shock proteins IbpA and IbpB and minor amounts the chaperones DnaK and GroEL[9,10,14,15]. Occasionally *E.coli* membrane components exist as contaminants because of unspecific attachment during the purification process[16]. Most commonly identified membrane proteins bound to IBs are plasmid encoded proteins and OmpT that are responsible for antibiotic resistance (kanamycin resistance protein and β-lactamase)[10,13].

Transmembrane proteins (TMPs) when expressed in heterologous systems such as *E.coli* have a natural tendency to aggregate into IBs. When expressed in *E.coli* membranes, they are still most subsequently extracted into detergents that are suitable for biophysical characterization. Thus in both cases, TMPs are extracted or are refolded into detergents, and in most cases activity of the protein is assessed based on ligand binding. If the activity of the protein is regained when refolded from native membranes or from IBs, the method is useful. Parameters such as quantities of expressed protein, the purification strategy, ease of isotope labeling, and other biochemical parameter are accounted in

identifying the best method. Thus from an NMR point of view, TMPs expressed as IBs in enhanced quantities and their reconstitution into detergent micelles when possible, is advantageous, because it allows the more efficient synthesis of isotope-labeled proteins.

Generally not all TMPs are expressed as IBs directly in *E.coli*. Reasons for directing a particular TMP into IB is not yet understood. When there is completely lack of expression of a particular type of TMP, insoluble fusion tags that direct proteins to IBs such as TrpΔLE, PurF, ketosteroid isomerase and PagP could help in TMP expression. Purifying IBs and reconstituting them into a membrane-type environment require special procedures, which are different from those used for soluble proteins. Protocols involving reconstitution take into account solubilizing the IBs in a strong denaturant such as Guanidinium Chloride (GdmHCl) or Urea, and then removing the denaturant through dilution, dialysis, or by chromatographic separation. This is followed by a refolding step into detergent micelles, bicelles or nanodiscs for NMR analysis.

In this chapter we discuss the production of various TM fragments of the neuropeptide Y4 receptor (NPY4R), which is a human GPCR (Fig.1) in the heterologous expression system *Escherichia coli*. Procedures involve direct expression or the usage of a fusion tag. Although we made attempts to use a soluble fusion tag such as SUMO for expressing the fragments, we were not successful in obtaining sufficient yields. When direct expression was possible, it was always directed into inclusion bodies. If the direct expression was not successful, we used the insoluble fusion tag, which is the TrpΔLE sequence that has already been discussed in the previous chapter. Basically, in the previous chapter we explained the constructs TM12, TM45 and TM67. In this chapter we focus on constructs of truncated TM67, TM56 with N- or C-terminal His-tag, TM567 and a Cys-free mutant of TM567.

Since all of the constructs were expressed as inclusion bodies both in LB (Luria Broth) and minimal medium, their purification strategy has been standardized to obtain high-quality protein for biophysical analysis. The standardized methodology explained in this chapter includes cloning, expression tests both in LB and minimal medium and the purification strategy.

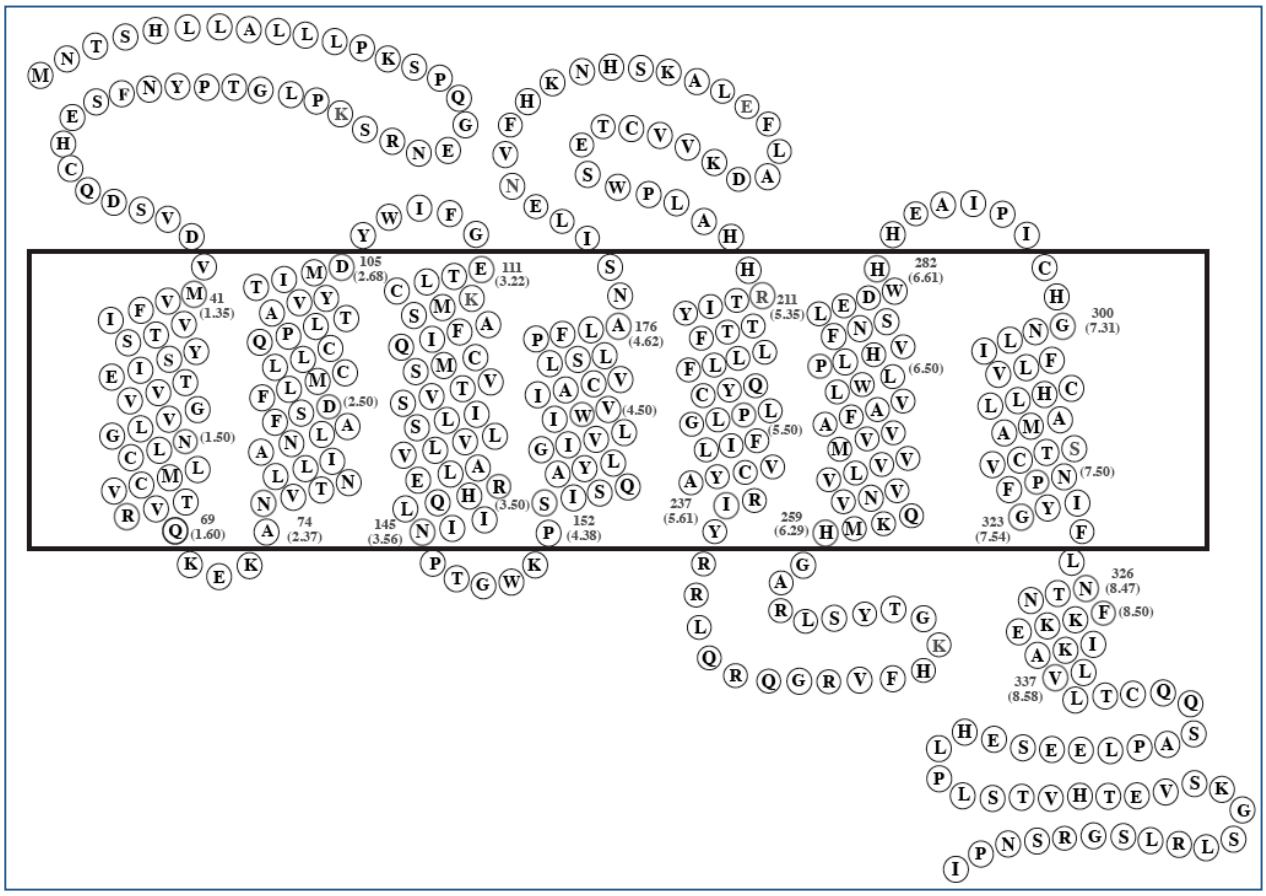


Figure 1. **Schematic representation of a human NPY4R**. The locations of extracellular, transmembrane and intracellular parts are depicted based on homology models of the receptor. Numbering refers to total receptor numbering (residue 1 refers to the N-terminus) or the Ballesteros-Weinstein nomenclature (in brackets).

3. BL21AI *E.coli* strain for expressing complex TMPs

E.coli BL21-type strains are mostly used because of useful features for producing eukaryotic TMPs. Some of the widely used BL21-type *E.coli* strains for recombinant protein expression are BL21(DE3), BL21(DE3)-pLysS, BL21 Star-pLysS, BL21AI, BL21 CodonPlus, C41(DE3) and C43(DE3). C41 and C43 *E.coli* strains are derivatives of BL21(DE3) strains. Although each strain has specific modification and has unique purpose, BL21AI stands out as the most suitable strain for us in expressing our TMP fragments.

Distinctive features of BL21AI strain compared to other *E.coli* strains are the lack of certain proteins such as *Ion* protease, outer membrane protease and OmpT. Absence of these proteases reduces the degradation of complex recombinant proteins. Additionally, the strain carries a chromosomal insertion of the gene encoding for T7 RNA polymerase (T7 RNAP) in the *ara*B locus of the *araBAD* operon, enabling expression of T7 RNAP to be controlled by the araBAD promoter (P_{BAD}) [17-19]. Thus T7 RNAP can be regulated by the sugar L-arabinose. Because T7 RNAP concentration is tightly regulated, the BL21AI strain is particularly useful in expressing recombinant proteins that may be toxic. In other *E.coli* strains that display leaky basal expression of T7 RNA polymerase (e.g. BL21 (DE3)) this would hamper cell growth.

4. Materials and methods

4.1 Cloning: The pLC01 plasmid was obtained from Prof. Fred Naider (College of Staten Island – CUNY, NY, USA). This plasmid originally contains a TrpΔLE gene cloned between Nde1 and HindIII restriction enzymes, and an Ampicillin resistance gene. The plasmid map of the pLC01 plasmid (without TrpΔLE) is presented in figure 2. Although the multiple cloning site contains more than 10 restriction recognition sites, the figure displays only 3 restriction sites that were used by us for most of our cloning purposes. The NPY4R fragment part of the gene that is to be inserted into pLC01 plasmid was amplified by polymerase chain reaction (PCR) from the wt-cDNA with the corresponding matching primers containing the restriction sites at the ends. Primers were ordered commercially from the supplier Microsynth AG.

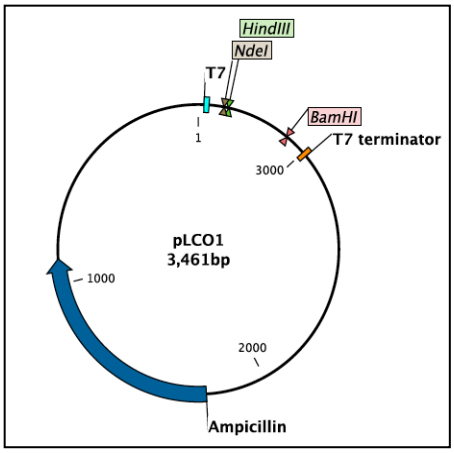
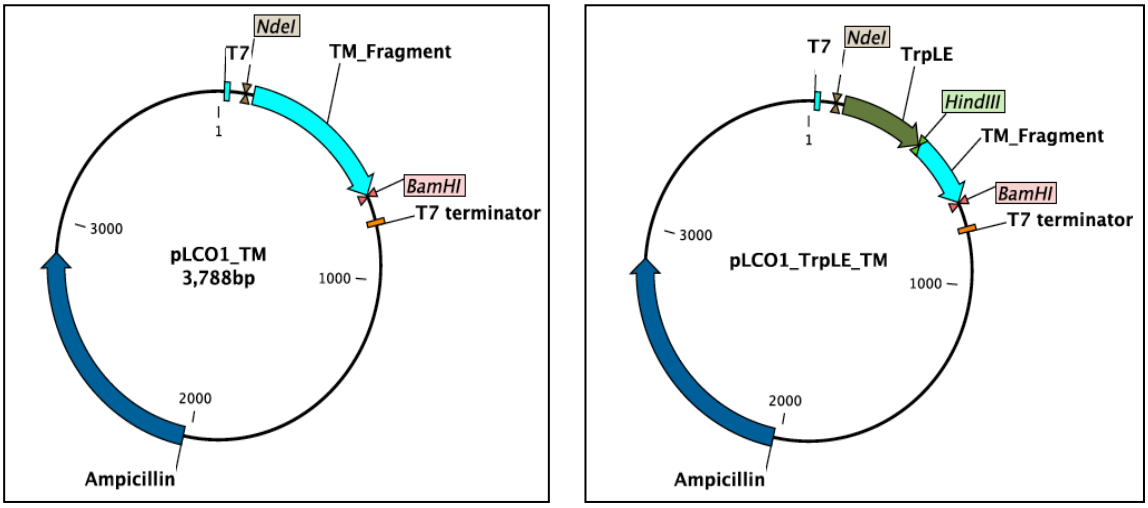


Figure 2. Schematic representation of the pLC01 plasmid containing the Ampicillin resistance gene and the restriction sites utilized for cloning.

For direct expressions, the gene of interest (TM fragment) is PCR amplified, purified and then double digested with Nde1 and BamH1, which subsequently is inserted into the plasmid pLC01 that is similarly doubly digested using the same restriction enzymes (figure3a). For introducing fusion tags such as SUMO or TrpΔLE, an overlapping PCR was performed between the fusion tag gene and the gene of interest (TM fragment) and then ligated into the Nde1 and BamH1 doubly digested plasmid. Alternatively, we generally maintained a pLC01 plasmid, which already contains the TrpΔLE fusion tag. In this case we just double digest the plasmid with HindIII and BamH1, and later ligated with the amplified TM gene of interest. For introducing the SUMO tag, we removed the TrpΔLE fusion gene by double digesting it with Nde1 and HindIII and later ligating it with the SUMO tag (figure 3b).



3a. **3b.**

Figure 3. Schematic representation of the pLC01 plasmid after introducing the TM fragment's gene of interest between Nde1 and BamH1 restriction sites (3a). In 3b is the representation of pLC01 plasmid after introducing the fusion tag between Nde1 and HindIII restriction sites and the TM fragment gene of interest between HindIII and BamH1 restriction sites.

For all the fragments of NPY4R, the target gene of interest was amplified from the wt-cDNA present originally in pcDNA 3.1, which is a mammalian expression vector. Almost all constructs are designed by introducing a hexa or a octa His-tag through cloning. For the Cys mutants, the constructs were generated from a synthetic gene of NPY4R devoid of Cys residues. The synthetic gene of NPY4R was commercially ordered from the company GeneArt. Sequences of all generated clones (table 1) were confirmed by sequencing. Gene and protein sequence analysis was done using CLC Main workbench software and EXPASY tools available online.

Sl.No.	Construct	Short Name	Primers (5'-3')
1.	His-NPY4R	HisY4	***Fwd:*** ATA CAT ATG CAT CAC CAT CAC CAT CAC AAC ACC TCT CAC CTC CTG GCC ***Rev***: AAT TAA GGA TCC TTA AAT GGG ATT GGA CCT GCC A
2.	NPY4R-His	Y4His	***Fwd:*** ATA CAT ATG AAC ACC TCT CAC CTC CTG GCC ***Rev***: ACG GGA TCC TTA ATG GTG ATG GTG ATG GTG ATG GTG AAT GGG ATT GGA CCT GCC
3.	ΔTrp-3C-w/o ICL3-TM67-w/o Cter	ΔxTM67x	***Fwd***: GAC AAG CTT CTG GAA GTG CTG TTC CAG GGG CCC ATG AAG CAG GTC AAT GTG GTG CTG ***Rev***: AAT TAA GGA TCC TTA AAT GGG ATT GGA CCT GCCA
4.	ΔTrp-3C-ICL3-TM67 w/o Cter	ΔTM67x	***Fwd*:** CTG GAA GTG CTG TTC CAG GGG CCC CGG CGC CTG CAG AGG CAG GG ***Rev***: TCT AGA CTC GGA TCC TTA CAG GGC CTT GAT CTC CTT CTT
5.	SUMO-3C-ICL3-TM67 w/o Cter	S-TM67-x	***Fwd:*** ATT CAT ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC GGC ***Rev*:** ATT AAG CTT ACC ACC AAT CTG TTC TCT GTG
6.	SUMO-3C-w/oICL3 -TM67-Cter	S-x-TM67	***Fwd:*** ATT CAT ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC GGC ***Rev*:** ATT AAG CTT ACC ACC AAT CTG TTC TCT GTG
7.	ΔTrp-3C-ECL2-TM56-ECL3	ΔTM56	***Fwd:*** GTG CTT TTC CAA GGC CCA GAG AAT GTC TTC CAC AAG AAC ***Rev*:** AGA CTC GGA TCC TTA GTT CCC ATG GCA GAT GGG GAT GGC
8.	His-ECL2-TM56-ECL3	HisTM56	***Fwd:*** AAT TTA CAT ATG CAT CAC CAT CAC CAT CAC CAT CAC ATC CTG GAG AAT GTC TTC ***Rev***: AAT TAA GGA TCC TTA GTT CCC GTG GCA GAT GGG GA

9.	ECL2-TM56-ECL3-His	TM56His	***Fwd:*** ATT CAT ATG ATC CTG GAG AATGTC TTC ***Rev*****:** CCT GGA TCC TTA ATG GTG ATG GTG ATG GTG GTT CCC ATG GCAGAT GGG
10.	ECL2-TM567-Cter-His	TM567His	***Fwd:*** ATT CAT ATG ATC CTG GAG AAT GTC TTC ***Rev***: ACG GGA TCC TTA ATG GTG ATG GTG ATG GTG ATG GTG AAT GGG ATT GGA CCT GCC
11*.	ECL2-TM567-ECL3-His (Cys free)	TM567His (Cys-free)	**C201S-*Fwd***: GCA GAT AAG GTG GAC AGT ACC GAG TCC TGG ***Rev***: CCA GGA CTC GGT ACT GTC CAC CTT ATC TGC
12*.	ΔTrp-3C-ICL3-TM67-Cter (Cysfree)	ΔTM67-Cys-free	**C298N-*Fwd***: GAG GCC ATC CCC ATC AAC CAT GGG AAC CTC ATC ***Rev***: GAT GAG GTT CCC ATG GTT GAT GGG GAT GGC CTC **C307F-*Fwd***: AAC CTC ATC TTC TTA GTG TTC CAC TTG CTT GCC ATG GCC ***Rev***: GGC CAT GGC AAG CAA GTG GAA CAC TAA GAA GAT GAG GTT **C316F-*Fwd***: CTT GCC ATG GCC TCC ACC TTT GTC AAC CCA TTC ATC TAT ***Rev***: ATA GAT GAA TGG GTT GAC AAA GGT GGA GGC CAT GGC AAG **C340K-*Fwd***: AAG GCC CTG GTG CTG ACT AAG CAG CAG AGC GCC CCC CTG ***Rev***: `CAG GGG GGC GCT CTG CTG CTT AGT CAG CAC CAG GGC CTT`

* These constructs/mutants were generated using the quick-change protocol as illustrated by the commercial supplier Stratagene.

Table 1. Different fragments of NPY4R generated by cloning into the pLC01 vector. Primers of each construct are listed.

4.2 Expression tests: After confirming the clones by sequencing, initial tests were done by transforming the plasmids into various *E.coli* expression strains. Protein overexpression was tested by small-scale screening, in which cells were grown in 1-3 ml of LB medium containing the appropriate antibiotic at 37^{o}C, and by using different inducer concentrations at different cell densities, and also by lowering temperatures after induction. In most cases, the BL21AI strain was used as the expression strain, requiring arabinose as its inducer. Best expression yields were obtained with 0.2% arabinose as the inducer at 20^{o}C in the BL21AI strain. Once overexpression was observed, LB medium was switched to minimal medium with ^{15}N-labeled NH_4Cl on a 1-liter scale. Based on the

protein overexpression levels per liter of medium, 1-3 liters of cultures were used for the subsequent purification steps. Basic ingredients in an M9 medium are listed below.

4.3 Minimal Medium Components

The following is the recipe for 1 liter of M9 minimal medium:

M9-salts:

- KH_2PO_4 ---------- 4 g/l
- K_2HPO_4 ---------- 4 g/l
- $Na_2HPO_4 * 2\ H_2O$ -- 3.5 g/l
- NaCl ---------- 1 g/l
- NH_4Cl ---------- 1 g/l

Note: Normal NH_4Cl was replaced with $^{15}NH_4Cl$ if ^{15}N labeling was desired.

To the above ingredients, 960 ml of deionized water was added to dissolve the salts and then the mixture was autoclaved. To the autoclaved 960 ml M9 salts solution, the following components were added to complete the M9 medium:

- 1 M $MgSO_4$ solution (autoclaved separately): 10 ml
- 20% (w/v) Glucose (filter sterilized): 25 ml, which gives final concentration of 5 g/l of M9 medium. For ^{13}C-labelling the final concentration of ^{13}C-Glucose was between 3-5 g/l.
- 150 mM Thiamine HCl (Vitamin B1; sterile filtered): 1 ml
- Antibiotic of choice: 1 ml
- Trace metal stock (sterile filtered): 1 ml
- pH of medium in the end was adjusted to 7.2

Trace metal stock composition:

$FeSO_4 * 7H_2O$ ---------- 4 g/l

$CaCl_2 * 2H_2O$ ---------- 4 g/l

$AlCl_3 * 6H_2O$ ---------- 1 g/l

$MnSO_4 * nH_2O$ ---------- 1 g/l

$CoCl_2 * 6H_2O$ ---------- 0.4 g/l

$ZnSO_4 * 7H_2O$ ---------- 0.2 g/l

$CuCl_2 * 2H_2O$ ---------- 0.1 g/l

H_3BO_3 ---------- 0.1 g/l

Caution: One must be careful while dissolving these salts, as some may react vigorously with water. Better is to dissolve them in a wide-open container (e.g. in a beaker) and not in a narrow vial (such as a Falcon tube). Some precipitate will be observed but this can be dissolved by drop wise addition of concentrated HCl. After the precipitate disappears/dissolves solution was sterile filtered.

4.4 Protein purification: The basic methodology for purifying proteins that are expressed in inclusion bodies was explained in the previous chapter. In brief, IBs isolated from sonicated cells were solubilized in 6 M GdmHCl, subjected to Ni-NTA purification and further purified on C4 RP-HPLC column. Wherever a fusion protein was involved, enzymatic cleavage was performed after affinity purification by adding the 3C precision protease to the elute obtained from affinity column. Subsequently, the denaturing buffer was removed by dialysis, resulting in precipitation of TM protein and fusion partner. The precipitate was re-solubilized in GdmHCl and further purified by C4 RP-HPLC using a water-acetonitrile gradient. All fragments of NPY4R overexpressed as IBs were subjected to the above standard purification method.

5. Results

5.1 Cell growth and protein expression: Once the cloned constructs were sequence confirmed, they were transformed into BL21AI *E.coli* cells for testing protein overexpression. The results of the expression tests are summarized in Table 2, and indicate that few of them showed very good and few absolutely no expression. Expression was initially screened on whole cells by looking at the cell lysates on SDS-page. Once expression occurs, it was investigated whether they accumulate in the soluble or in the insoluble fractions. In almost all cases the overexpressed protein was observed in the inclusion bodies.

Sl.No.	Construct	Short Name	Fusion	Expression	Can be purified
1.	His-NPY4R	HisY4	----	No	------
2.	NPY4R-His	Y4His	----	No	------
3.	ΔTrp-3C-w/o ICL3-TM67-w/o Cter	ΔxTM67x	TrpΔLE	Yes	No
4.	ΔTrp-3C-ICL3-TM67 w/o Cter	ΔTM67x	TrpΔLE	Yes	No
5.	SUMO-3C-ICL3-TM67 w/o Cter	S-TM67-x	SUMO	No	-----
6.	SUMO-3C-w/oICL3 -TM67-Cter	S-x-TM67	SUMO	No	-----
7.	ΔTrp-3C-ECL2-TM56-ECL3	ΔTM56	TrpΔLE	Yes	No
8.	His-ECL2-TM56-ECL3	HisTM56	----	Yes	Yes
9.	ECL2-TM56-ECL3-His	TM56His	----	Yes (No in D_2O)	Yes
10.	ECL2-TM567-Cter-His	TM567His	----	Yes	Yes
11.	ECL2-TM567-ECL3-His (Cys free)	TM567His -(Cys free)	----	Yes	Yes
12.	ΔTrp-3C-ICL3-TM67-Cter (Cysfree)	ΔTM67-Cysfree	TrpΔLE	Yes	Yes

Table 2. Expression yields and success of purification for different fragments of NPY4R.

Note: As the discussion progresses, I would be using the symbol 'Δ' in denoting ΔTrp and the symbol 'x' in denoting the removal of N- or C-terminus.

Expression of the full-length receptor with either N- or C-terminal His-tag was not possible. Although few of them such as ΔTM67x and ΔxTM67x were expressed as IBs

with the help of a fusion-tag, the enzymatic cleavage and its subsequent purification proved to be very difficult. To bypass the difficulties related to poor solubility we also used the soluble SUMO fusion instead, but this resulted in a total loss of protein expression. A new construct of ΔTM56 was made with the fusion to TrpΔLE, which gave identical problems as in the case of ΔTM67x and ΔxTM67x. Subsequently, direct expressions were tested for TM56 with a His-tag on either N- or C-terminus, which both (TM56His and HisTM56) gave positive results with respect to expression and purification. Finally, TM7 was added to the TM56 construct to test direct expression of longer fragments. Surprisingly the new 3-TM construct TM567His displayed even better expression than the 2-TM constructs of TM56 (HisTM56 & TM56His). Since the 3-TM construct was also a direct expression in the form of IBs, no cleavage was required and purification was straightforward.

In the following table the yields are summarized for proteins that were subjected to more detailed biophysical analysis (table 3).

Sl.No.	**Construct**	**LB (mg/l of culture)**	**M9-H_2O (mg/l)**	**M9-D_2O (mg/l)**
1.	HisTM56	2	4	0.75
2.	TM56His	1	1.5-2	0
3.	TM567His	3	10	0.8
4.	TM567His (Cys-free)	4	10	*
5.	ΔTM67 (Cys-free)	3	5	*

* Not tested yet

Table 3. Protein yields of different NPY4R fragments in 3 different media.

It is also interesting to note the growth parameters such as pH and the cell-densities of BL21AI *E.coli* strain under different labeling conditions at the start and the end of cell cultures (table 4). In general it is assumed that the cells stop growing if they are unable to express the recombinant proteins, but I have observed the contrary with the BL21AI cells as they always were growing irrespective of the TMP was expressed or not. Growth parameters indicated in the below table was a standard phenomenon for the BL21AI strain for all expressions attempted on the C-terminal NPY4R fragments.

Factor		LB	M9-H_2O	M9-D_2O
pH	Start	7.2	7.2	7.2
	End	6.4	5.8	6.8
Time required to reach induction OD_{600} of 0.8		2hrs	5-6hrs	20-24hrs
Maximum cell density (OD_{600}) reached after 24hrs of induction at 20^oC (48hrs in case of M9-D_2O)		1.8	4.2	1.8

Table 4. Factors fluctuating while growing BL21AI strain in different media.

It is interesting to note that the yields are comparably high in M9-H_2O, which might be correlated to high cell densities as observed in table 4. Another aspect is the drastic reduction of protein yields in M9-D_2O, which might be correlated to extremely slow growth rate as observed from the induction times in table 4.

If we consider the availability of various factors/nutrients for the growth of BL21AI *E.coli* strain in LB and M9 minimal medium, cells were not able to grow beyond a cell density of 1.8 in LB but to high densities of above 4 in M9-H_2O. Changing from H_2O to D_2O with the same M9 components, growth had been reduced drastically from 4.2 to 1.8 even after considering their slow nature of growth.

5.2 Protein Purification: Since all fragments expressed into IBs, they were solubilized using 6M GdmHCl containing Tris buffer (20 mM Tris - pH 8.0, 150 mM NaCl and 10 mM β-mercaptoethanol) for overnight. Solubilized IBs were loaded onto a Ni-NTA column pre-equilibrated with the above denaturing buffer, washed sequentially each time with three column volumes of the same buffer additionally containing 10 mM, 20 mM and 30mM imidazole. His-tagged protein was eluted with 200 mM Imidazole containing denaturing buffer, which was further purified using C4 RP-HPLC column (VYDAC). Before loading into the C4 column, pH of the Ni-NTA elute was reduced to 2, which allows binding to the hydrophobic column which is pre-equilibrated with the solvent mixture – 90% water and 10 Acetonitrile (both contain 0.1% TFA). Acetonitrile concentration was gradually increased allowing the column bound protein to be

completely soluble at certain water-acetonitrile gradient. The gradient concentration, at which the protein separates from the column, depends on its hydrophobicity.

Attempts were made to purify all the expressed proteins. Fragments such as ΔTM67x, ΔxTM67x, ΔTM56, HisTM56, TM56His, TM567His, TM567His (Cys-free) and ΔTM67 (Cys-free) of NPY4R had shown expression and thus we will discuss the purification of each fragment of NPY4R separately. Each fragment will be denoted with a snakeplot-type schematic representation, exact sequence taken from the protoparam file format and their chromatographic purification followed by analysis on SDS-PAGE.

5.2.1 Construct 1. ΔTrp-3C-w/o ICL3-TM67-w/o Cter (ΔxTM67x):

xTM67x Sequence after cleavage:

```
         10        20        30        40        50        60
GPMKQVNVVL VVMVVAFAVL WLPLHVFNSL EDWHHEAIPI CHGNLIFLVC HLLAMASTCV

         70
NPFIYGFLNT NFKKEIKAL
```

Residues: 79; Molecular weight: 8.91 kDa

From the previous chapter, NMR characterization of a TM67 fragment containing ICL3 and the C-terminal domain was near to 80%, but the assignments of the TM regions (TM6 & TM7) were difficult because peaks were either weak or absent. Most of TM6 was only partially assigned. The reasons might be attributed to the redundancy of valines in the sequence (VNVVLVVMVVA) resulting in peak overlap or extensive line broadening due to conformational exchange effects. To at least overcome the peak overlap problem, the following construct was designed by removing the complete ICL3 loop and nearly 70% of C-terminus while keeping the soluble helix 8 intact. This construct was well expressed as a fusion in the IBs but the enzymatic cleavage of the fusion partner was reduced by 50% compared to the original construct that contained the full loops (fig 4c and fig 4d). Although a small peak of xTM67x was observed at 74%

water- acetonitrile gradient, SDS-PAGE analysis showed large contamination from the uncleaved fusion.

Additional feature of this construct was its very high hydrophobicity, which might not be suitable to work with C4 RP-HPLC purification. I observed that there is considerable amount of protein loss on injecting the cleaved products into the hydrophobic column and alternatively the xTM67X product would very easily aggregate with its partners rendering the purification process altogether difficult. Thus this fragment was not analyzed further primarily because of the very low efficiency of the enzymatic cleavage of the fusion and its difficulty to handle in solubilizing it in monodisperse form.

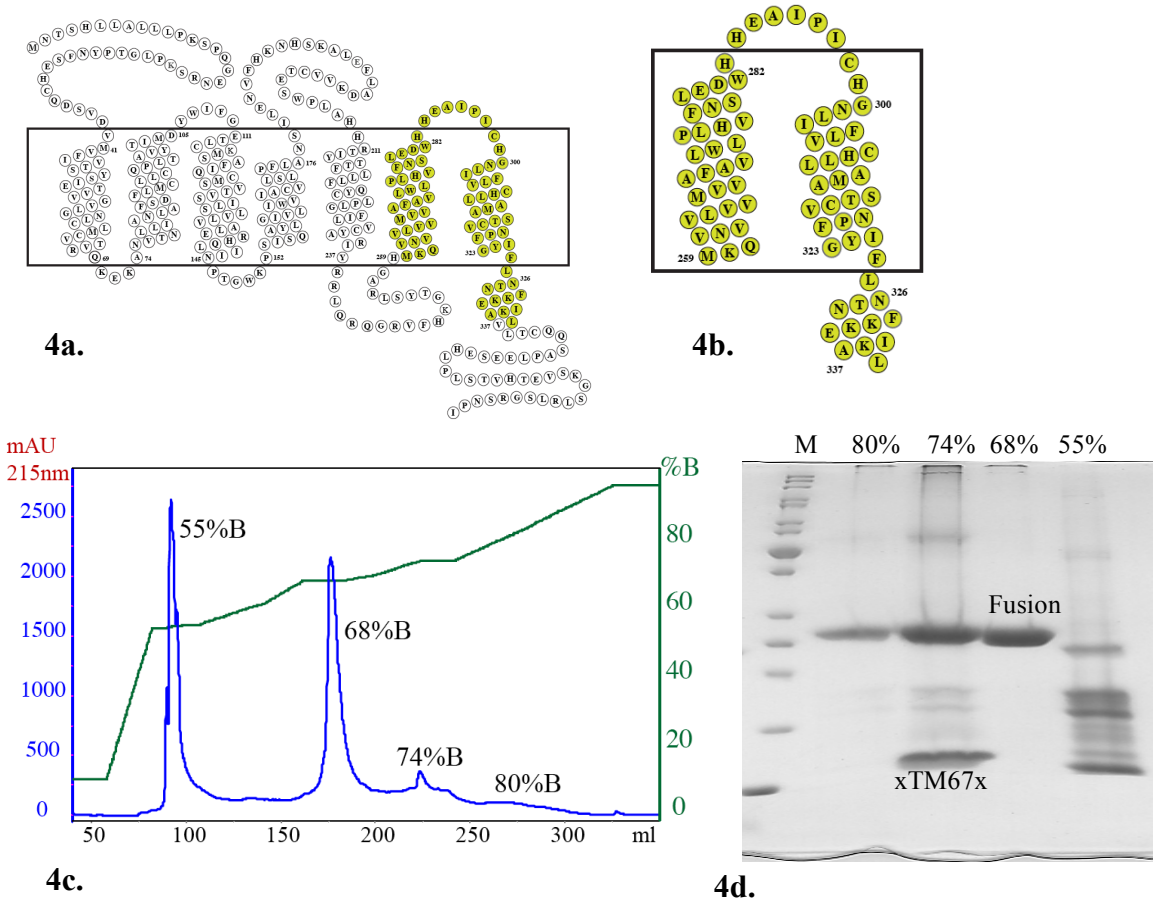
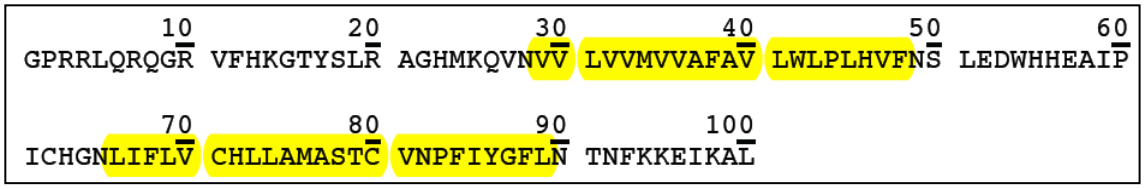


Figure 4. Snake plot of the Y4 receptor with the ΔxTM67x region highlighted in yellow (4a and 4b). RP-HPLC chromatogram of ΔxTM67x after performing the precision protease cleavage (4c) and the purity of each observed peak was analyzed by SDS-PAGE (4d).

5.2.2 Construct 2. ΔTrp-3C-ICL3-TM67 w/o Cter (ΔTM67x):

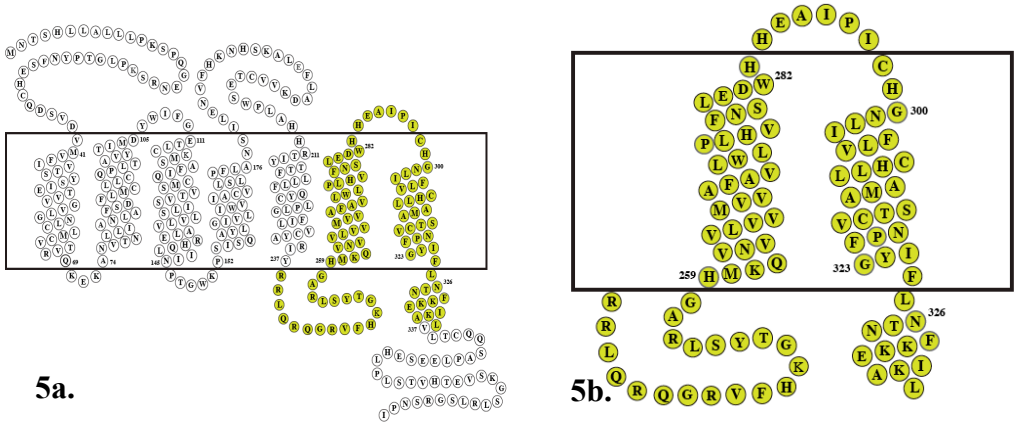
TM67x Sequence after cleavage:



Residues: 100; Molecular weight: 11.42kD

We reasoned that the cleavage site might be not accessible because it is too close to the hydrophobic TM segments. Accordingly, ICL3 was re-introduced (fig 5a and 5b). When tested, the enzymatic cleavage improved compared to ΔxTM67x, but the TM67x fragment obtained from the cleavage of fusion construct (ΔTM67x) tends to aggregate on the column as the peak of interest was of very low intensity while eluting with the water – acetonitrile gradient (fig 5c). We observe strong peak for the ΔTrp which normally elutes at 55% of water-acetonitrile gradient and thus conclude that this construct was good for C3 protease cleavage but not for hydrophobic purification. The cleaved TM67x product tends to be very hydrophobic that is resulting in strong interactions the hydrophobic tails of the column or could be co-aggregating and precipitating in the column.

We speculate that the long C-terminus (nearly 45 residues) is necessary to keep the fragment in the monomeric state. Although the cleavage yields are better than for the previous construct, they were not as reproducible as in the original TM67 fragment discussed in the previous chapter.



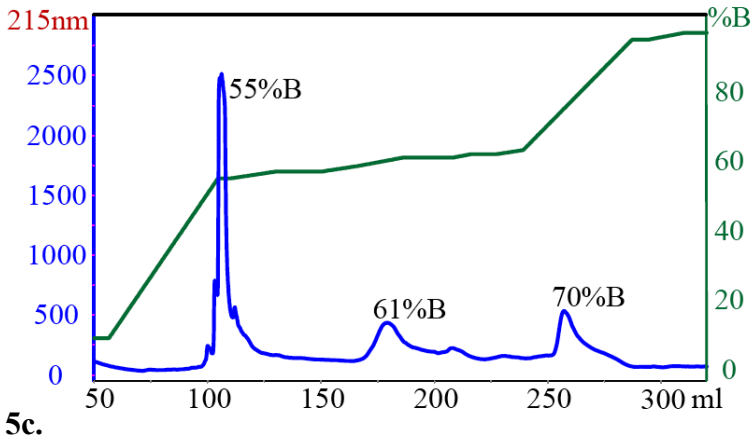


Figure 5. Snake plot of the Y4 receptor with the TM67x region highlighted in yellow (5a and 5b). Purification of TM67x fragment by reverse-phase HPLC after performing the precision protease cleavage (5c). Since the peaks obtained were of very low intensity, I didn't analyze the fractions by SDS-PAGE.

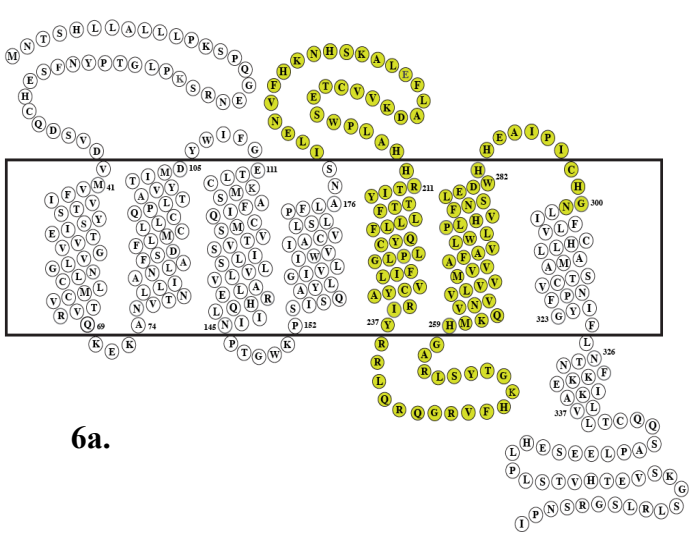
5.2.3 Construct 3. ΔTrp-3C-ECL2-TM56-ECL3 (ΔTM56):

TM56 Sequence after cleavage:

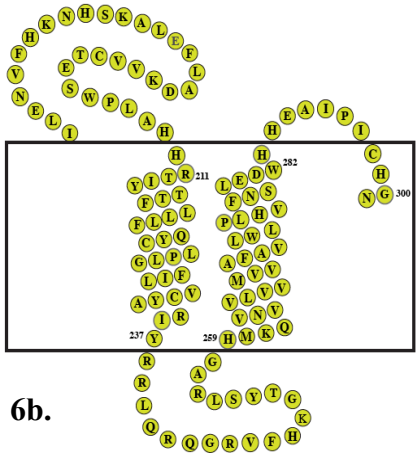
```
        10         20         30         40         50         60
GPENVFHKNH SKALEFLADK VVCTESWPLA HHRTIYTTFL LLFQYCLPLG FILVCYARIY

        70         80         90        100        110        120
RRLQRQGRVF HKGTYSLRAG HMKQVNVVLV VMVVAFAVLW LPLHVFNSLE DWHHEAIPIC

HGN
```

Residues: 123; Molecular weight: 14.32 kDa

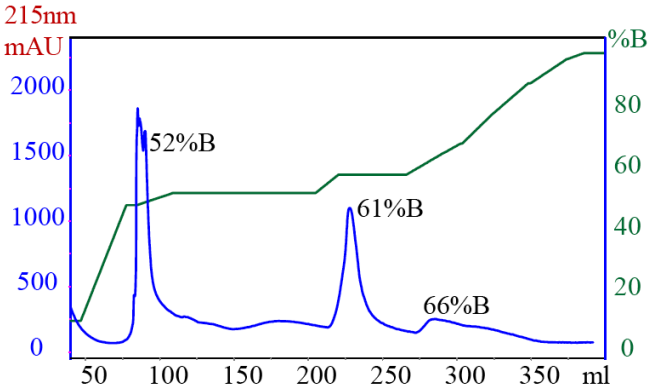
Similar to construct 1, the idea behind designing this construct was to better understand the reason for the observed line-broadening of resonances from TM6. We suspected that the broadening might be due to interactions with TM5 that take place on an unfavorable timescale. Thus TM56 was designed as a fusion to TrpΔLE (fig 6a and 6b). As a fusion, the fragment ΔTM56 was well expressed but the cleavage efficiency was so low that the purified TM56 could not be identified on the SDS-PAGE (fig 6c and 6d).



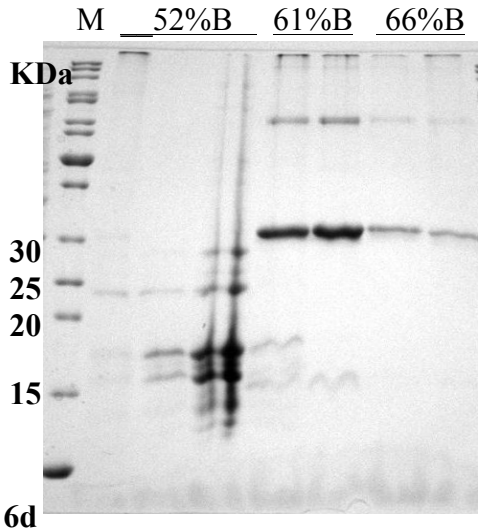
6a.



6b.



6c



6d

Figure 6. Snake plot of the Y4 receptor with the TM56 region highlighted in yellow (6a and 6b). Purification of TM56 fragment by reverse-phase HPLC (6c) after cleavage and SDS-PAGE of the fractions from the HPLC (6d).

From the chromatogram (6c) and the SDS-Page gel (6d) analysis, the peak at 52% water-acetonitrile gradient was supposed to be the cleaved TrpΔLE, the peak at 61% acetonitrile gradient was the uncleaved fusion and the broad peak at 66% gradient also shows the uncleaved fusion on SDS-PAGE due to the tailing of prior peak at 61%. I assume that the peak at 66%B contains TM56 at very low concentrations, which can't be detected using SDS-PAGE. Since the cleavage was low and the purification didn't show promiscuous outlook for more optimizations, I decided to check for direct expression of the same construct.

5.2.4 Construct 4. His-ECL2-TM56-ECL3 (HisTM56):

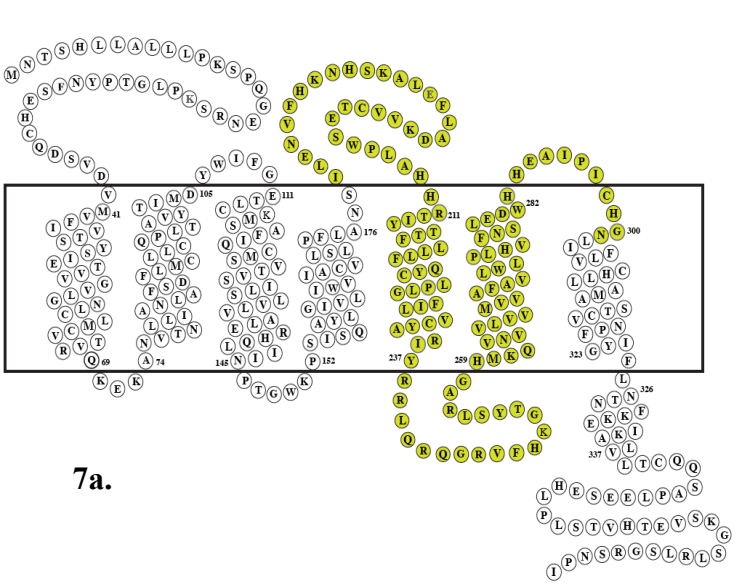
HisTM56 Sequence:

```
         10         20         30         40         50         60
MHHHHHHHHI LENVFHKNHS KALEFLADKV VCTESWPLAH HRTIYTTFLL LFQYCLPLGF

         70         80         90        100        110        120
ILVCYARIYR RLQRQGRVFH KGTYSLRAGH MKQVNVVLVV MVVAFAVLWL PLHVFNSLED

        130
WHHEAIPICH GN
```

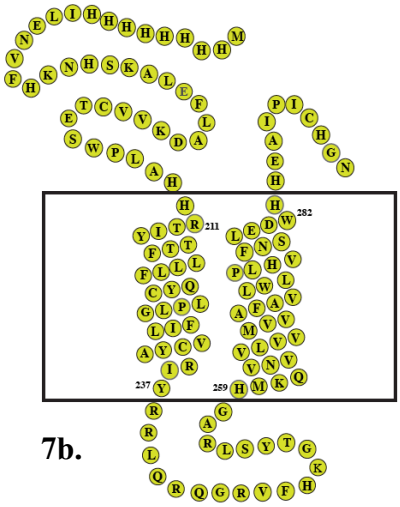
Residues: 132; Molecular weight: 15.62 kDa

To avoid the problems that occurred during cleavage and purification of the fusion proteins, HisTM56 was designed for direct expression. An octa-His-tag was introduced at the N-terminus of TM56 (fig 7a and 7b) for affinity purification. Following expression HisTM56 was observed in the inclusion bodies and a purification protocol similar to that used for the previous constructs was followed, except for the need for enzymatic cleavage. After affinity purification, HisTM56 was loaded onto a C4 column for RP-HPLC purification and the purified fragment eluted at 62% acetonitrile/water gradient (fig 7c), which is confirmed by SDS-PAGE analysis (fig 7d). As discussed in the table 3, this construct gave sufficient yields (4 mg/l in M9-H_2O) for generating an NMR sample and for spectroscopic characterization. This fragment was expressed in very low quantities when changed from M9-H_2O to M9-D_2O (0.75 mg/l) and the yields were down to only 20% of M9-H_2O expression. NMR characterization was performed with this fragment, which will be discussed in the following chapter.

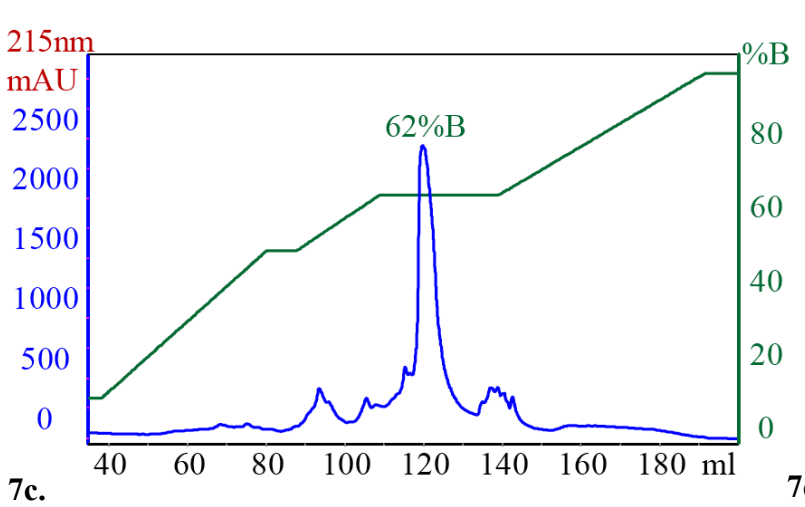
This was my first construct where I was successful in identifying and purifying a NPY4R fragment without the usage of a fusion tag. Although reasonable yields were obtained from IBs, I also tested if any fraction of the protein goes into the membranes. I collected the supernatant or the soluble fraction from the sonicated cells and tried to extract the His-tag containing proteins using Ni-NTA resin. On SDS-PAGE analysis, I wasn't able to see any protein at the respective molecular weight of TM56 indicating all of the expressed TM56 was diverted into IBs.



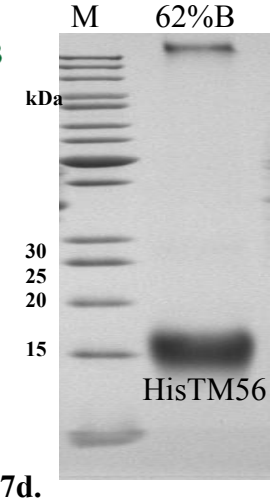
7a.



7b.



7c.



7d.

Figure 7. Snake plot of the Y4 receptor with the HisTM56 region highlighted in yellow (7a and 7b). HPLC chromatogram of HisTM56 (7c) and SDS-PAGE of the main peak (7d).

5.2.5 Construct 5. ECL2-TM56-ECL3-His (TM56His):

TM56His sequence:

```
         10          20          30          40          50          60
MILENVFHKN HSKALEFLAD KVVCTESWPL AHHRTIYTTF LLLFQYCLPL GFILVCYART

         70          80          90         100         110         120
YRRLQRQGRV FHKGTYSLRA GHMKQVNVVL VVMVVAFAVL WLPLHVFNSL EDWHHEAIPI

        130
CHGNHHHHHH
```

Residues: 130; Molecular weight: 15.35 kDa

Since few of the TM residues in the previous construct were unable to be analyzed (assigned) by NMR, We speculated an interaction with the N-terminal His-tag resulted in poor spectral quality of the previous construct, and thus we designed a new construct of identical sequence but with the His-tag (6x) attached to the C-terminus. Expression yields were reduced to <50% compared to the previous construct (table 3), pointing to an important role of the position of the His-tag. While sufficient yields were obtained for biophysical analysis in M9-H_2O (fig 8c and 8d) no expression was observed in M9-D_2O. Exhaustive screening by varying temperature after induction, inducer concentration and inducing cell densities was done for the expression of TM56His in D_2O but with no success. Probably this is related to the fact that expression yields were already low in M9-H_2O compared to the previous construct.

As indicated in table 3, this construct yielded less than half of the quantities (around 1.5 – 2 mg/l) in M9-H_2O compared to the prior construct and to make things worse, it hasn't expressed at all in M9-D_2O. At this stage, not knowing the facts that direct expressions yields drop drastically in M9-D_2O, I made many despairing attempts changing conditions (temperature and induction parameters) to check for expression on a small scale. Considering TM7 in TM67 behaves well (TM7 backbone was nearly completely

assigned), I attempted to add TM7 to TM56 generating new construct with a hope to obtain better yields and thus moved on to the next construct TM567His.

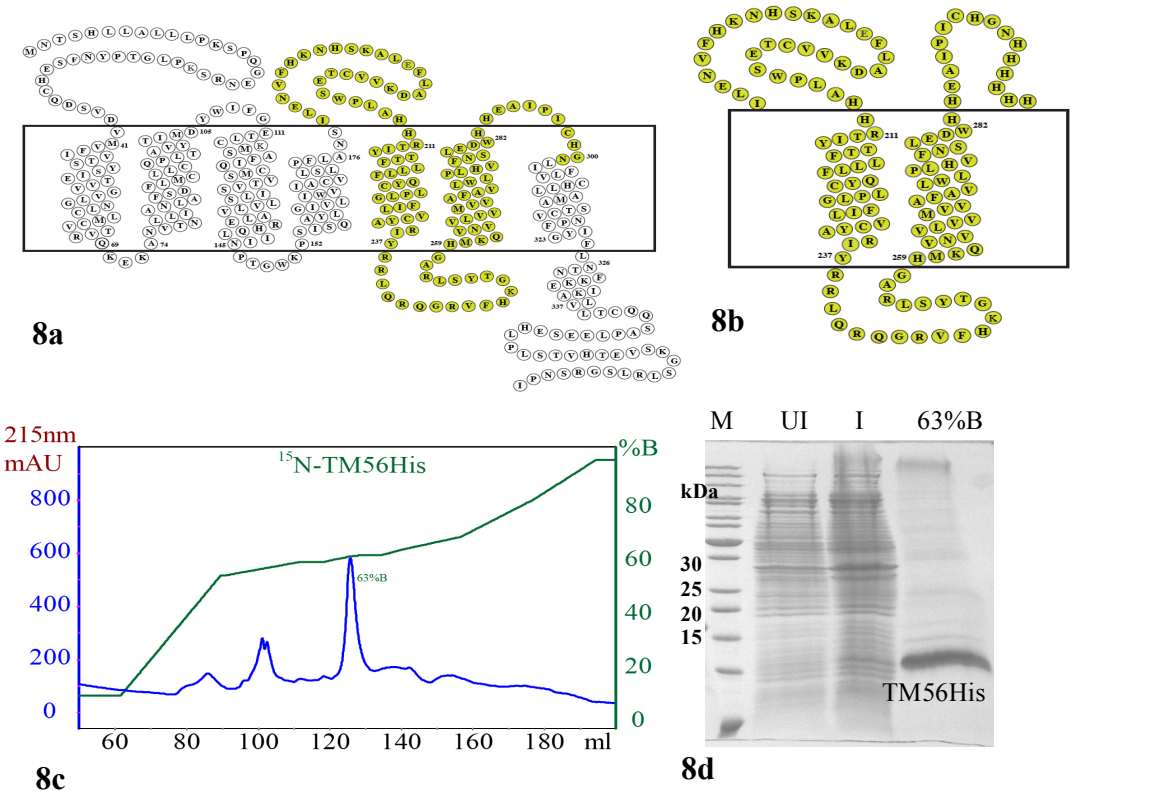


Figure 8. Snake plot of the Y4 receptor with the TM56His region highlighted in yellow (8a and 8b). HPLC Chromatogram of ^{15}N-TM56His (8c). The fraction at 63% was analyzed by SDS-PAGE (8d). UI – Uninduced; I – Induced.

5.2.6 Construct 6. ECL2-TM567-Cter-His (TM567His):

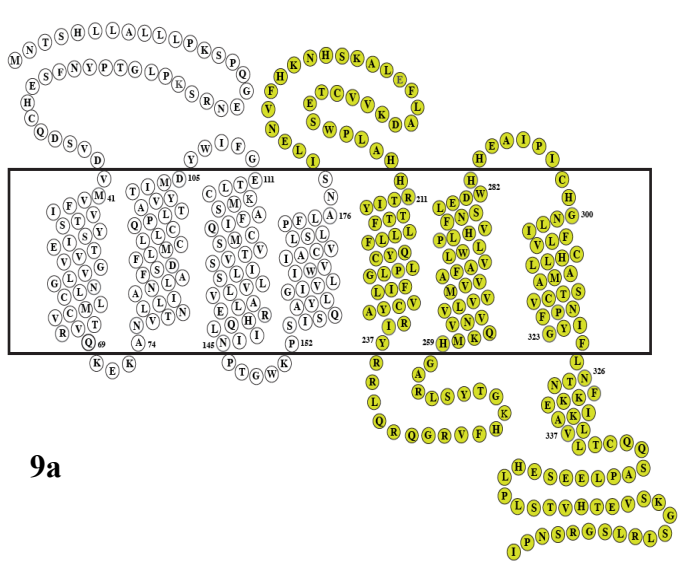
TM567His Sequence:

```
        10         20         30         40         50         60
MILENVFHKN HSKALEFLAD KVVCTESWPL AHHRTIYTTF LLLFQYCLPL GFILVCYARI

        70         80         90        100        110        120
YRRLQRQGRV FHKGTYSLRA GHMKQVNVVL VVMVVAFAVL WLPLHVFNSL EDWHHEAIPI

       130        140        150        160        170        180
CHGNLIFLVC HLLAMASTCV NPFIYGFLNT NFKKEIKALV LTCQQSAPLE ESEHLPLSTV

       190        200
HTEVSKGSLR LSGRSNPIHH HHHHHH
```

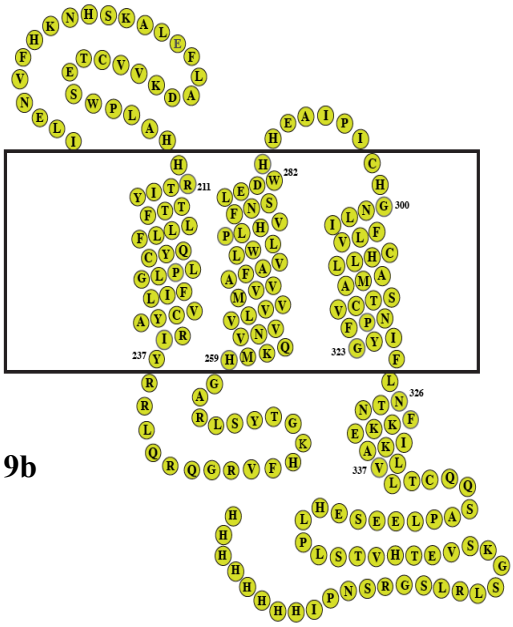
Residues: 206; Molecular weight: 23.76 kDa

One of the project's goal was to investigate to which extend chemical shifts can be transferred from the shorter to the longer fragments. Moreover, we were interested to learn how additional helices may stabilize secondary or tertiary structure, and whether they improve integration into the membrane. To this end the following fragment TM567His was designed with an octa-His-tag attached to the C-terminus for direct expression (fig 9a and 9b).

Considering the low expression yields of the previous fragment (TM56His), we assumed the larger fragment to be expressed in extremely low yields. Surprisingly, this fragment expressed much better than any of the previously described fragments. Protein was directed into inclusion bodies, purified by Ni-NTA in denaturing buffer (GdmHCl) and subsequently refined by C4 RP-HPLC (fig 9c). The fragment TM567His eluted at 64% acetonitrile gradient in a pure form, which was analyzed by SDS-PAGE (fig 9d).



9a



9b

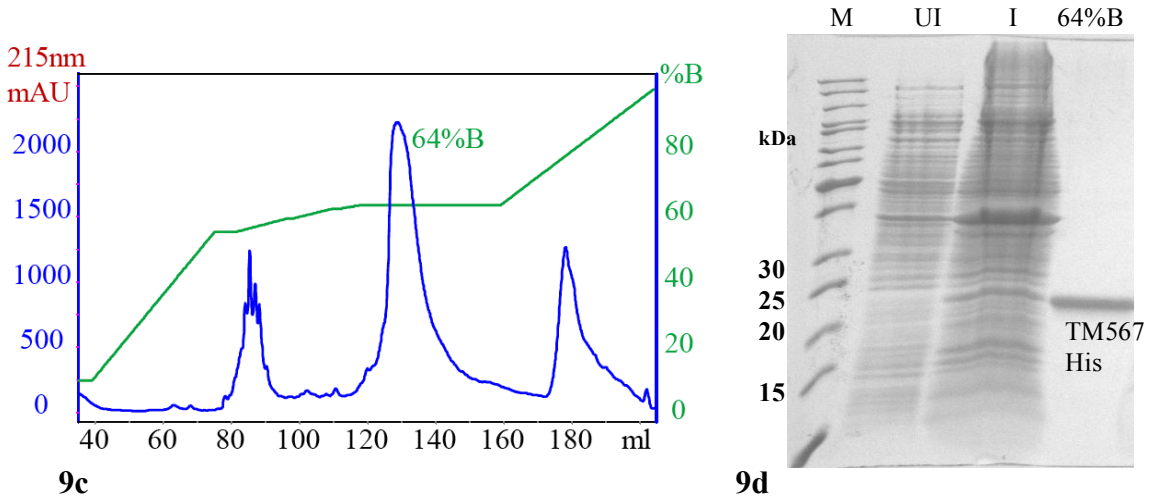


Figure 9. Snake plot of the Y4 receptor with the TM567His region highlighted in yellow (9a and 9b). HPLC chromatogram of TM567His (9c). SDS-PAGE of the 64% water-acetonitrile fraction as well as with cell-lysates (9d). UI – Uninduced; I – Induced.

5.2.7 Construct 7. ECL2-TM567-Cter-His w/o Cys (TM567His Cys-free):

TM567His-Cysfree Sequence:

```
         10         20         30         40         50         60
MILENVFHKN HSKALEFLAD KVDSTESWPL AHHRTIYTTF LLLFQYVLPL GFILVSYARI

         70         80         90        100        110        120
YRRLQRQGRV FHKGTYSLRA GHMKQVNVVL VVMVVAFAVL WLPLHVFNSL EDWHHEAIPI

        130        140        150        160        170        180
NHGNLIFLVF HLLAMASTFV NPFIYGFLNT NFKKEIKALV LTKQQSAPLE ESEHLPLSTV

        190        200
HTEVSKGSLR LSGRSNPIHH HHHHHH
```

Residues: 206; Molecular weight: 23.87 kDa

Broad lines in the previous construct lead us to suspect that aggregation occurred through formation of intermolecular disulfide bonds. Accordingly, we designed a construct, which is free of Cys amino acids. Moreover, a Cys-free mutant would be required for attaching MTSL spin labels to unique sites to allow for PRE experiments. The mutated residues are underlined in red in the above sequence, and the nature of the replacement was decided based on a sequence alignment of NPY receptor homologues. The most conserved residues were selected as Cys replacement. Thus we ordered a synthetic gene

of NPY4R, which is completely free of Cys residues. From the synthetic gene, the TM567 construct was generated by adding a His-tag (8x) to the C-terminal end.

Expression was followed similar to the wild type fragment, which was purified initially using affinity chromatography and then followed by C4 RP-HPLC purification (10c). The major peak, which is the purified TM567His Cys-free protein, was confirmed by SDS-PAGE analysis (10d).

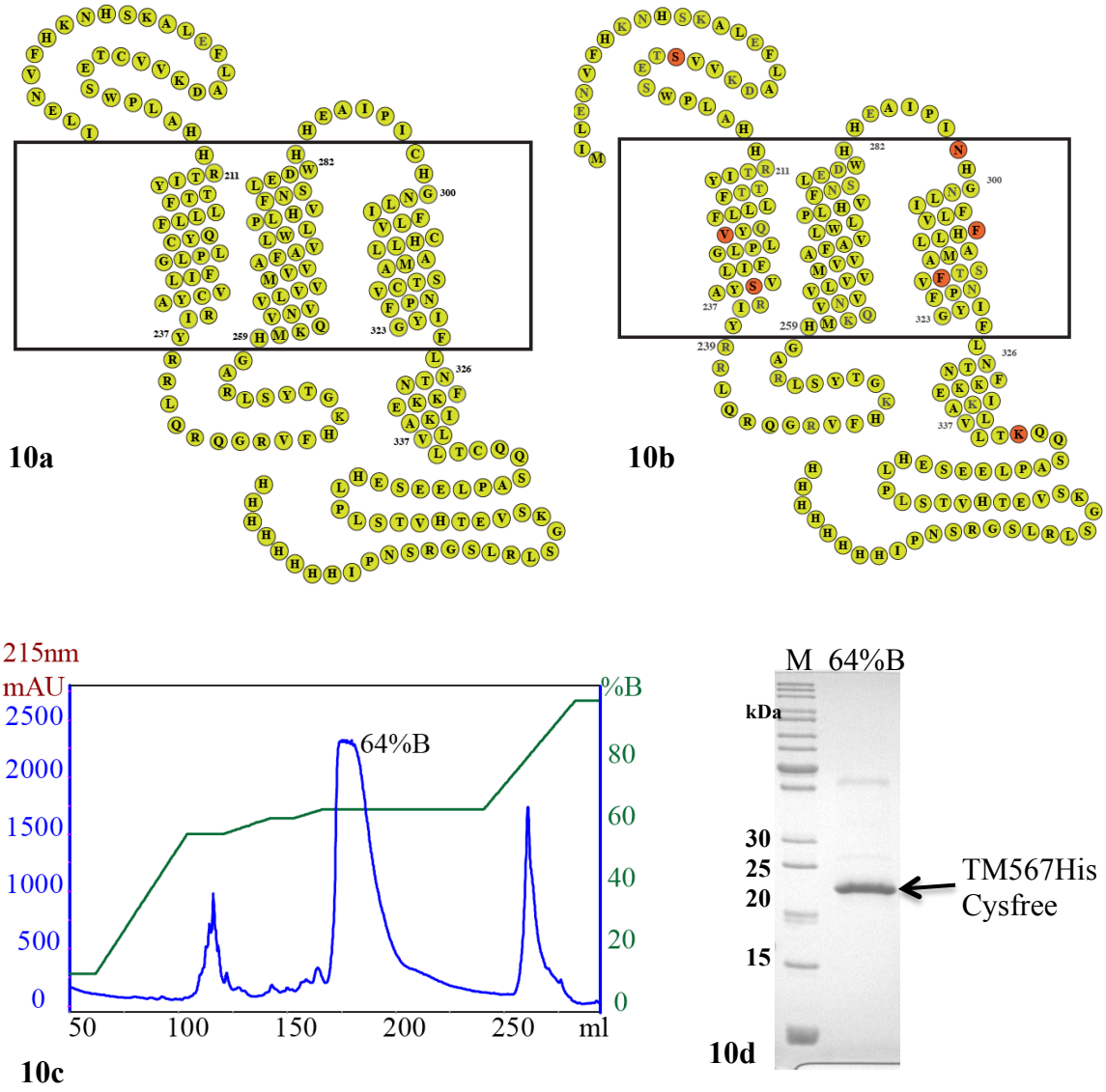


Figure 10. Snake plot of the wild type TM567His (10a) and of TM567His-Cysfree mutant (10b) with the replaced Cys-residues highlighted in red. HPLC chromatogram of TM567His (10c) and SDS-PAGE of the main HPLC peak at 64% ACN (10d).

6. Discussion

Excluding rhodopsin, all GPCRs exist in their native membranes at moderately low levels, and thus for structural investigations the only practical way in obtaining them in milligram quantities is by recombinant expression in heterologous hosts. Heterologous hosts used for GPCR expression are bacteria, yeast, mammalian cells, insect cells infected by baculoviruses, and using host extracts for cell-free synthesis. Recently, most of the GPCR structures analyzed were obtained by expressing them in insect cells, which provide milligram quantities of protein for crystallization purposes[20-25]. However, host systems such as mammalian- or insect cells cannot be adapted for the synthesizing isotope labeled proteins required for NMR because of the complexity and high costs associated with the nutrients involved.

Eukaryotic transmembrane proteins can sometimes be expressed in very simple to handle heterologous systems such as bacteria, but post-translational modifications and proper refolding remains as an ongoing issue. However, many trials proved that the detergent reconstituted TMPs from bacterial membranes[26-28] or refolded proteins from IBs[29-33] have native folding mimicking the activity when present in native environments. In general, expression of heterologous proteins in bacterial strains is not generally possible. Protein expression levels depend on amino acid sequence, the type of bacterial strain as well as on details of expression conditions (temperature, availability of protein folding machinery and their sustaining environments). Many a times these factors could be circumvented by using fusion tags that help in folding, directing to bacterial membranes or to inclusion bodies. Usage of a fusion tag could give good protein yields or good quality protein but the purification processes involving cleavage are most of the times complicated resulting in unnecessary time-consumption. To understand the nature of widely used fusion tags, we shall examine the N-terminal amino acid sequence of first 30 residues.

	Protein	**Aminoacid sequence of soluble fusion tags**
1.	Ubiquitin	MQIFV KTLTG KTITL EVEPS DTIEN VKAKI
2.	SUMO	MSDQE AKPST EDLGD KKEGE YIKLK VIGQD

3.	MBP	MKIEE GKLVI WINGD KGYNG LAEVG KKFEK

Table 5. First 30 residues of the most commonly used soluble fusion-tags. Charged aminoacids are highlighted in red; polar aminoacids in blue and the remaining are the hydrophobic aminoacids.

These soluble fusion tags are dominated by charged and polar residues, which could also help in solubilizing or proper folding of certain proteins. On the other hand fusion tags (few examples shown below), which direct the proteins into insoluble IBs, contain equal amounts of hydrophobic residues spread over the sequence.

	Protein	**Aminoacid sequence of insoluble fusion tags**
4.	TrpΔLE	MKAIF VLKGS LDRDL DSRIE LELRT DHKEL
5.	KSI	MREYS SRQTR RGCSP PCRRI EAKVI DGRNY
6.	PagP	MNVSK YVAIF SFVFI QLISV GKVFA NADEW

Table 6. First 30 residues of the most commonly used insoluble fusion-tags.

Sometimes a TMP is expressed in the *E.coli* expression system even in absence of a fusion tag, and will most of the time be observed in the insoluble inclusion body fraction. Reasons for few of the protein expressions directed to membranes and few of them into inclusion bodies are again not known. But it is interesting to compare the N-terminal sequences of our fragments with those of both the widely used soluble and insoluble fusion protein sequences. Another interesting fact observed in NPY4R fragments is the expression of TM56His in very low quantities compared to TM567His, which is contrary to the fact that in *E.coli* expression of shorter TMPs should be easier than larger multi-helical TMPs.

	Protein	**Aminoacid sequence of NPY4R fragment constructs**	**Yields**
7.	TM67	RRLQR QGRVF HKGTY SLRAG HMKQV NVVLV	0
8.	HisTM56	MHHHH HHHHI LENVF HKNHS KALEF LADKV	4 mg/l
9.	TM56His	MILEN VFHKN HSKAL EFLAD KVVCT ESWPL	2 mg/l
10.	TM567His	MILEN VFHKN HSKAL EFLAD. ……+TM7	10 mg/l

Table 7. First 30 residues of different NPY4R fragments used in the study.

A comparison amino acid occurrence of all the sequences mentioned above is shown in table 8. A few interesting points can be noted such as

- Arg, Cys and His residues are not observed at the beginning of soluble fusion-tags.
- Met if present is not found at any other position than the first one.
- Presence of a His-tag at the N-terminus improves expression as observed from the fragments HisTM56 and TM56His.

Amino acid	1 Ubq	2 SUMO	3 MBP	4 ΔTrp	5 KSI	6 PagP	7 TM67	8 His-TM56	9 TM56-His	10 TM567-His	Natural occurrence (%)
R	0	0	0	3	7	0	5	0	0	0	5.4
D	1	4	1	4	1	1	0	1	1	1	5.4
E	3	4	4	3	2	1	0	2	3	3	6.6
H	0	0	0	1	0	0	2	10	2	2	2.4
K	4	5	6	3	1	2	2	3	3	3	6.1
Charged											25.9
C	0	0	0	0	2	0	0	0	1	1	1.7
N	1	0	2	0	1	2	1	2	2	2	4.9
Q	1	2	0	0	1	1	3	0	0	0	4.2
S	1	2	0	2	3	3	1	1	2	2	8.4
T	5	1	0	1	1	0	1	0	1	1	5.6
Y	0	1	1	0	2	1	1	0	0	0	3.1
Polar											27.9
G	1	3	5	1	2	1	3	0	0	0	6.0
A	1	1	1	1	1	3	1	2	2	2	6.9
I	4	2	3	2	2	3	0	1	1	1	5.6
L	2	2	3	6	0	1	3	3	4	4	9.3
M	1	1	1	1	1	1	1	1	1	1	2.2
F	1	0	1	1	0	4	1	2	2	2	4.0
P	1	1	0	0	2	0	0	0	1	1	5.2
W	0	0	1	0	0	1	0	0	1	1	1.2
V	3	1	2	1	1	5	5	2	3	3	6.0
Others											46.4

Table 8. Comparison of the first few (30) N-terminal residues in commonly expressed soluble proteins, IBs and NPY4R fragments. Charged aminoacids are highlighted in red;

polar aminoacids in blue and the remaining are the hydrophobic aminoacids. The last column displays the frequency of each aminoacid occurrence within a protein sequence as observed in eukaryotes[34].

In this chapter we have also established a generic purification method, starting from GdmHCl-solubilized inclusion bodies. We note that the enzymatic cleavage must be improved, e.g. by fine-tuning the dialysis rate that competes with the enzymatic cleavage, varying the concentration of GdmHCl to have a good compromise of enzyme activity, determined by the concentration at which the precision protease folds with full activity and the point at which the enzymatic cleavage site becomes accessible to the refolded enzyme.

7. References

1. Jurgen, B. et al. Quality control of inclusion bodies in Escherichia coli. *Microb Cell Fact* **9**, 41 (2010).
2. Wang, L. Towards revealing the structure of bacterial inclusion bodies. *Prion* **3**, 139-145 (2009).
3. Wasmer, C. et al. Solid-state NMR spectroscopy reveals that E. coli inclusion bodies of HET-s(218-289) are amyloids. *Angew Chem Int Ed Engl* **48**, 4858-4860 (2009).
4. Ventura, S. Sequence determinants of protein aggregation: tools to increase protein solubility. *Microb Cell Fact* **4**, 11 (2005).
5. Carrio, M., Gonzalez-Montalban, N., Vera, A., Villaverde, A. & Ventura, S. Amyloid-like properties of bacterial inclusion bodies. *J Mol Biol* **347**, 1025-1037 (2005).
6. Bowden, G. A., Paredes, A. M. & Georgiou, G. Structure and morphology of protein inclusion bodies in Escherichia coli. *Biotechnology (N Y)* **9**, 725-730 (1991).
7. Marston, F. A. The purification of eukaryotic polypeptides synthesized in Escherichia coli. *Biochem J* **240**, 1-12 (1986).
8. Villaverde, A. & Carrio, M. M. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnol Lett* **25**, 1385-1395 (2003).
9. Carrio, M. M. & Villaverde, A. Construction and deconstruction of bacterial inclusion bodies. *J Biotechnol* **96**, 3-12 (2002).
10. Jurgen, B. et al. Monitoring of genes that respond to overproduction of an insoluble recombinant protein in Escherichia coli glucose-limited fed-batch fermentations. *Biotechnol Bioeng* **70**, 217-224 (2000).
11. Valax, P. & Georgiou, G. Molecular characterization of beta-lactamase inclusion bodies produced in Escherichia coli. 1. Composition. *Biotechnol Prog* **9**, 539-547

(1993).
12. Rinas, U., Boone, T. C. & Bailey, J. E. Characterization of inclusion bodies in recombinant Escherichia coli producing high levels of porcine somatotropin. *J Biotechnol* **28**, 313-320 (1993).
13. Rinas, U. & Bailey, J. E. Protein compositional analysis of inclusion bodies produced in recombinant Escherichia coli. *Appl Microbiol Biotechnol* **37**, 609-614 (1992).
14. Hoffmann, F. & Rinas, U. Kinetics of heat-shock response and inclusion body formation during temperature-induced production of basic fibroblast growth factor in high-cell-density cultures of recombinant Escherichia coli. *Biotechnol Prog* **16**, 1000-1007 (2000).
15. Allen, S. P., Polazzi, J. O., Gierse, J. K. & Easton, A. M. Two novel heat shock genes encoding proteins produced in response to heterologous protein expression in Escherichia coli. *J Bacteriol* **174**, 6938-6947 (1992).
16. Georgiou, G. & Valax, P. Isolating inclusion bodies from bacteria. *Methods Enzymol* **309**, 48-58 (1999).
17. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113-130 (1986).
18. Lee, D. H., Huo, L. & Schleif, R. Repression of the araBAD promoter from araO1. *J Mol Biol* **224**, 335-341 (1992).
19. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**, 4121-4130 (1995).
20. Massotte, D. G protein-coupled receptor overexpression with the baculovirus-insect cell system: a tool for structural and functional studies. *Biochim Biophys Acta* **1610**, 77-89 (2003).
21. Mancia, F. & Hendrickson, W. A. Expression of recombinant G-protein coupled receptors for structural biology. *Mol Biosyst* **3**, 723-734 (2007).
22. Rasmussen, S. G. et al. Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549-555 (2011).
23. Chiu, M. et al. Heterologous production of active mammalian G protein-coupled receptors using baculovirus-infected insect cells. In: Production of membrane proteins (Robinson, A. S, editor., Ed.) 109-138 (2011).
24. Manglik, A. et al. Crystal structure of the micro-opioid receptor bound to a morphinan antagonist. *Nature* **485**, 321-326 (2012).
25. Granier, S. et al. Structure of the delta-opioid receptor bound to naltrindole. *Nature* **485**, 400-404 (2012).
26. Sarkar, C. A. et al. Directed evolution of a G protein-coupled receptor for expression, stability, and binding selectivity. *Proc Natl Acad Sci U S A* **105**, 14808-14813 (2008).
27. Dodevski, I. & Pluckthun, A. Evolution of three human GPCRs for higher expression and stability. *J Mol Biol* **408**, 599-615 (2011).
28. Egloff, P. et al. Structure of signaling-competent neurotensin receptor 1 obtained by directed evolution in Escherichia coli. *Proc Natl Acad Sci U S A* **111**, E655-E662 (2014).
29. Kiefer, H. et al. Expression of an olfactory receptor in Escherichia coli: purification, reconstitution, and ligand binding. *Biochemistry* **35**, 16077-16084 (1996).

30. Rogl, H., Kosemund, K., Kuhlbrandt, W. & Collinson, I. Refolding of Escherichia coli produced membrane protein inclusion bodies immobilised by nickel chelating chromatography. *FEBS Lett* **432**, 21-26 (1998).
31. Ma, C. et al. Expression, purification, and activities of full-length and truncated versions of the integral membrane protein Vpu from HIV-1. *Protein Sci* **11**, 546-557 (2002).
32. Kiefer, H. In vitro folding of alpha-helical membrane proteins. *Biochim Biophys Acta* **1610**, 57-62 (2003).
33. Indiveri, C., Galluccio, M., Scalise, M. & Pochini, L. Strategies of bacterial over expression of membrane transporters relevant in human health: the successful case of the three members of OCTN subfamily. *Mol Biotechnol* **54**, 724-736 (2013).
34. Tekaia, F. & Yeramian, E. Evolution of proteomes: fundamental signatures and global trends in amino acid compositions. *BMC Genomics* **7**, 307 (2006).

Chapter IV: Biophysical Characterization of the C-terminal NPY4R Fragments by NMR

1. Abstract

Structural characterization of membrane proteins is often dependent on the availability of sufficient amounts of purified protein and the type of membrane mimetic used. Until now, the best method for structure determination of TMPs has been X-ray crystallography, but not all TMPs (especially a few G-Protein Coupled Receptors - GPCRs) are amenable for structure elucidation by this method. Inherent characteristics of GPCRs such as their hydrophobic nature make the expression and purification difficult. Their flexibility, and the requirement for membrane-like environments to insure proper folding are limiting steps in structure determination. Also, NMR is more suitable to study dynamics, conformational equilibria and ligand interactions than X-ray crystallography.

In the last 2-3 decades, solution NMR spectroscopy technique has evolved and diversified significantly from its applications for small molecules to biological macromolecules, and is widely used for structure determination of proteins, DNA and RNA. Nowadays, NMR is persistently being used even on molecules amenable to crystallography because of its tremendous advantages for functional investigations that include protein dynamics, conformational equilibria and ligand interactions. Application of NMR spectroscopy on TMPs is again challenging, and presents a very active field of research advancing with new methods. The main difficulties when using NMR on TMPs are large sizes of TMPs (taking into account the membrane environment), intrinsic dynamic properties and the low sensitivity of the technique.

In this chapter, we discuss details of the NMR sample preparation of all the C-terminal NPY4R fragments and the related assignment strategies. I also discuss the usage of chemical shifts in transferring assignments onto larger fragments based on NOESY peaks. I conclude the chapter discussing possible reasons for the missing peaks and the peaks, which could not be assigned.

2. Introduction

Solution NMR has been used for almost three decades to gain insights into the structural, conformational and interactive dynamics of TMPs in the presence of non-native membranes. Major developments on NMR instruments have taken place only in the last decade by the emergence of high-strength magnets, which improve the signal dispersion as well as the signal-noise ratio. These advancements helped scientists to transfer their expertise from small peptides to larger proteins. But the ability to replicate the same proficiency from even small TM peptides to small TMPs is until today though exciting and stimulating still challenging, arduous and costly with very few successful exceptions[1-9]. The main advantage of working with NMR over crystallography on TMPs is the diversity of environments that it can support: from native cell-membrane environment (in solid-state NMR) to different membrane mimetic environments such as micelles[4,7], bicelles[10], planar bilayers[1], nanodiscs[11], and organic solvents[12] at different salt concentrations, temperatures, and pH values. These days, solution NMR technique is also being used to study proteins in intact cells by in-cell NMR[13,14]. Adding to the advantages of NMR spectroscopy over crystallography, NMR is an indispensible tool in drug screening for crucial proteins involved in diseases[15,16].

Recently, a number of novel NMR methods have been developed, and in combination with specialized labeling patterns for proteins have helped to gain assignments. Even though much advancement has been made, an NMR study of TMPs present in their native membranes is only possible if large amounts (mg) of the protein can be expressed and purified and this is true only in very few cases like bacteriorhodopsin. Solution NMR techniques applied today on TMPs involved basically the methodologies originally developed for large soluble proteins. These techniques were applied not only for TMP structure determination but also for dynamics, and interactions with partners[17]. There are very few reviews describing the techniques and practical challenges involved in the application of NMR spectroscopy of TMPs[18-20]. Despite the difficulties related to the preparation of protein samples and the interpretation of the complicated spectra that mostly are of poor quality, some studies focused on structural characterization of TMPs by NMR have been reported, providing useful knowledge in structure-based drug discovery[21,22].

In the last decade, studies on several intact GPCRs have significantly advanced because of the technical developments witnessed in crystallizing the proteins especially in lipidic cubic phases (LCP)[23-25]. Many techniques are under development for crystallizing intact GPCRs and there still exist many challenges to study them at atomic resolution. We herein propose an alternative approach, in which folding and structure of these receptors is understood by using smaller segments, allowing the transfer of knowledge onto larger fragments, and ultimately onto the full intact protein. Working with smaller fragments will help in establishing and standardizing experimental procedures required to investigate more complicated systems, especially concerning protein-detergent or protein-lipid interactions. In most cases, production of intact GPCRs results in limited quantities of purified protein, which restricts the scope of screening diverse factors like membrane mimetics, ligand interactions, pH and temperature. Thus, to overcome the above limitation, we decided to understand the behavior of various NPY4R fragments comprising 2-3 TM in size and transfer their knowledge onto larger or full-length proteins.

In this chapter we report the NMR spectroscopic analysis of three C-terminal fragments of NPY4R - TM67, TM56 and Cys-free TM567His mutant (including wild-type). Synthesis of all fragments was explained in detail in the previous chapter. NMR spectra acquisition and backbone analysis of TM67 was explained in the second chapter and published in the ChemBioChem journal[26]. Here, we report the progress on the side chain assignment of TM67 and the backbone assignments of TM56 and the Cys-free variant of TM567 in presence of the detergent 1-palmitoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)] (LPPG).

3. Materials and methods

3.1 Chemicals

All chemicals and solutions such as $^{15}NH_4Cl$, ^{13}C-D-Glucose, ^{13}C-d_7-D-Glucose and D_2O were bought from Spectra Stable Isotopes (Andover, MA). 1-palmitoyl-2- hydroxy-*sn*-glycero-3-[phospho-rac-(1-glycerol)] (LPPG) was purchased from Anatrace products

(Maumee, OH). Remaining chemicals were acquired from Sigma-Aldrich (Buchs, Switzerland).

3.2. Sample preparation and NMR Spectroscopy

Biophysical characterization of TMPs or fragments of TMPs by solution NMR spectroscopy primarily involves production and purification of ^{15}N labeled protein, which has been discussed in detail in the previous chapter. Purified fragments were reconstituted into different detergent micelles, allowing collection of 2D ^{1}H-^{15}N correlation NMR spectra, which were initially analyzed with respect to signal to noise, chemical shift dispersion, line-widths and the number of observed peaks. In our experience LPPG displays the most favorable properties for multi-TM GPCR fragments, resulting in spectra with better signal dispersion and more homogenous line-widths[26-28]. After identifying the most suitable detergent and the appropriate conditions for NMR measurements, a 3D NOESY-HSQC spectrum was recorded. Although also sequential NOEs are found in these spectra most peaks stem from intra-residual correlations helping to identify the nature of the amino acid. Subsequently, a uniformly labeled ^{15}N, ^{13}C and ^{2}H protein was expressed and purified. Sample was prepared using the optimized conditions, which allowed measuring a series of standard 2D and 3D triple-resonance experiments, followed by respective peak picking of correlated resonances. NMR experiments measured and required for protein backbone assignment are TROSY, HNCO, HNCACO, HNCA, HNCOCA, HNCACB and HNCOCACB. All triple-resonance spectra relied on TROSY-type magnetization transfer except those required for side-chain assignment. With the help of strip matching, assignment of backbone resonances specific to the aminoacid sequence was achieved. Attempts for the identification of the missing assignments were made either by designing new constructs or by using alternative NMR experiments that involved shorter coherence transfer pathways, e.g. using the HNCA instead of the HNCACB. Once backbone assignment was at a point when no more correlations were found, a uniformly labeled ^{15}N and ^{13}C protein was expressed and purified, and spectra for side-chain assignments were recorded in deuterated d_{36}-LPPG.

All NMR measurements were done at 320 K on a Bruker AV700 spectrometer equipped with a triple-resonance cryoprobe. NMR samples contained approximately 0.4 mM protein in 40 mM phosphate buffer, pH = 5.4 with 40 mM DTT and 5% LPPG as described in chapter 2[26]. Proton chemical shifts were referenced using the water line at 4.48 ppm at 320 K.

3.3 Construct 1. ICL3-TM6-ECL3-TM7-Cter (TM67):

The cloning, expression and backbone assignment of TM67 was explained in chapter 2[26]. The final status of the backbone assignment is represented in figure 1 and figure 2 on the following page. Almost 80% of the sequence was assigned unambiguously and the missing assignments were mostly from TM6. We speculated that some weak signals (ambiguous assignments) of the missing residues may be attributed to Val repeats in the TM6 sequence (**V**N**VV**L**VV**M**VV**A), and missing peaks to extensive line broadening or peak overlap. To overcome some of these effects and to assign the missing residues, we attempted to design and purify shorter fragments eliminating the long N- and C-terminal tails. Although we were able to express the fusion constructs, their purification attempts were unsuccessful likely due to the more hydrophobic nature of these proteins. The biochemistry of the shorter versions of TM67 has been explained in the previous chapter.

Since we had more than 80% backbone assignment of TM67, we progressed to obtain its side-chain assignment. For this purpose, we expressed and purified a uniformly labeled ^{15}N and ^{13}C fusion protein and purified the TM67 fragment as described in the previous chapter[26]. Similarly, we prepared a 0.4 mM NMR sample solubilized in phosphate buffer containing 5% d_{36}-LPPG detergent. This deuterated detergent helps in eliminating the intense residual signals emerging from the detergent that would otherwise obscure the C-α and methyl resonances of aminoacid residues in the protein.

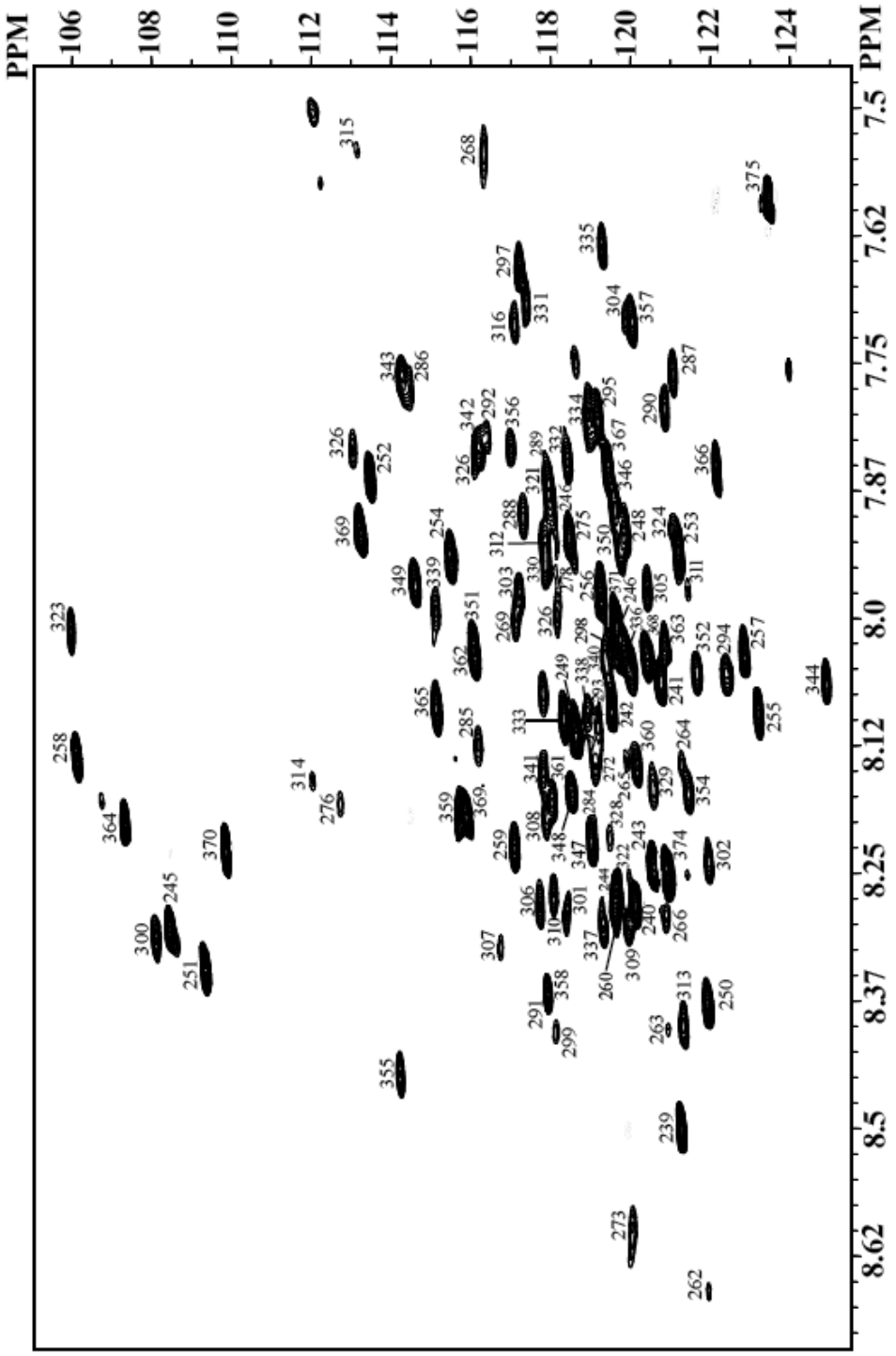


Figure 1: [^{15}N,^{1}H]-TROSY spectrum of TM67 with the assigned resonances. Numbers indicated on top of each resonance represents the corresponding aminoacid residue present in the full-length NPY4R.

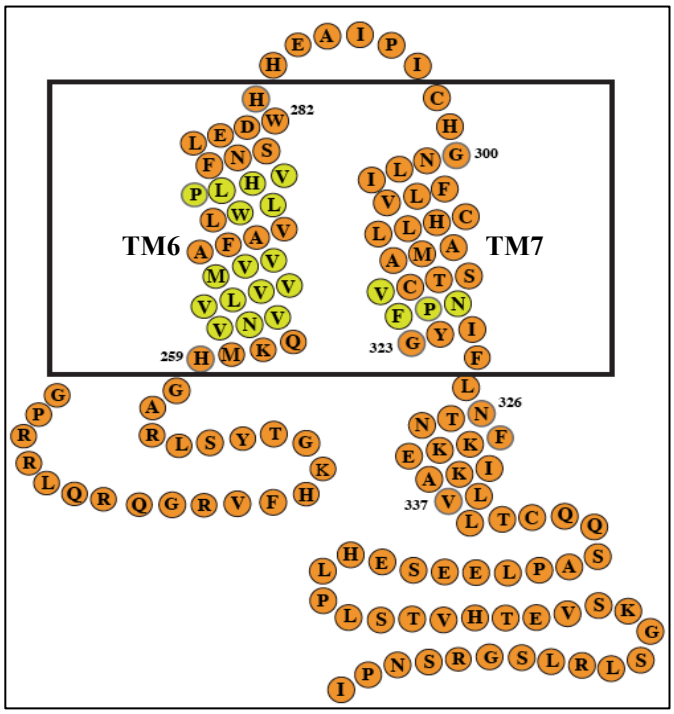


Figure 2: Snakeplot of TM67 with the completed assignments in orange and the missing assignments in green. Most of the unassigned region falls in TM6.

Unfortunately we noticed significant shifts of peak positions in [^{15}N, ^{1}H]-HSQC spectra, along with a large inhomogeneity in line-widths between spectra measured using deuterated and non-deuterated detergent and between deuterated and non-deuterated proteins (figure 3).

All the NMR data acquisition parameters are summarized in table 1. All spectra were initially processed using the Bruker spectrometer software Topspin 2.1 and the chemical shift assignment was performed using the CARA software tool (Keller 2004).

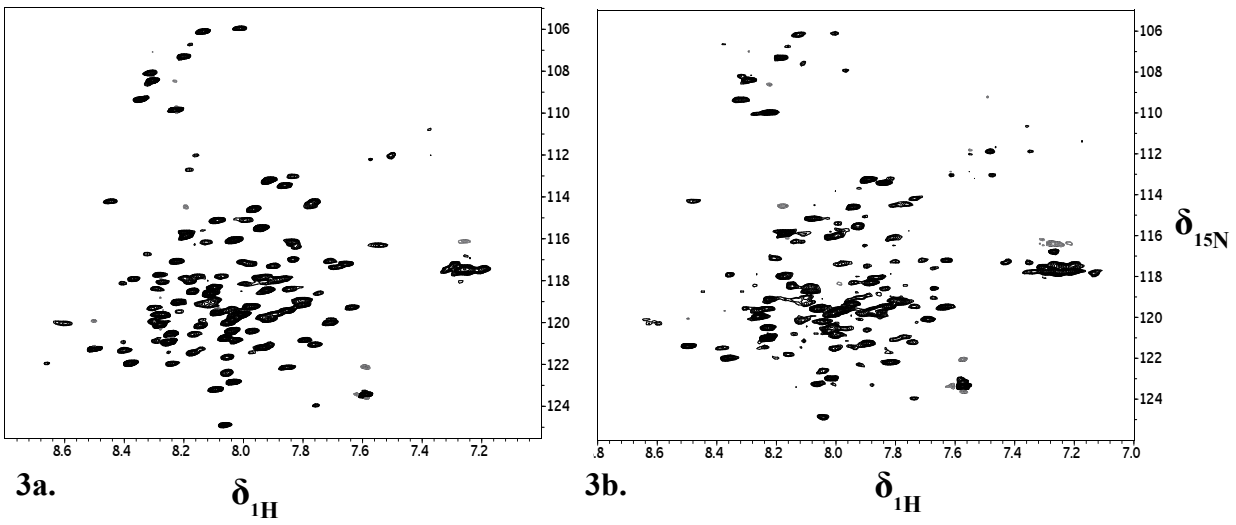


Figure 3: [^{15}N, ^{1}H]-TROSY spectra of TM67 when labeled differently. Spectrum (3a) was used for backbone assignment on a uniformly labeled ^{15}N, ^{13}C and ^{2}H protein dissolved in protonated LPPG and spectrum (3b) was used for side-chain assignment on a uniformly labeled ^{15}N and ^{13}C protein dissolved in deuterated LPPG.

TM67								
Experiment	# of data points			Spectral Width (ppm)			Scans	Remarks
	F3	F2	F1	F3	F2	F1		
[^{15}N,^{1}H]-HSQC		2048	256		18	33	32	
HNCO	2048	40	128	18	22	22	8	
HNCACO	2048	40	110	18	22	22	16	
HNCACB	2048	40	124	18	22	70	32	
HN(CO)CACB	2048	40	128	18	22	70	16	
HNCA	2048	40	92	18	22	20	32	
HN(CO)CA	2048	40	97	18	22	18	32	
^{15}N-NOESY	2048	40	161	14	22	10	16	120ms mix
[^{13}C,^{1}H]-HSQC		2048	256		13	70	32	
HCCH-TOCSY	2048	48	120	13	25	70	32	
^{13}Cali-NOESY	2048	40	120	13	33	10	24	100ms mix

Table 1: Spectroscopic details of NMR experiments used for measuring TM67

Before starting the peak-picking from the ^{13}C-edited spectra, Cα and Cβ chemical shifts generated from the backbone assignments were first adjusted using hCCH-TOCSY spectrum to correct for the isotope shifts. Chemical shifts were ultimately correlated to

the peak positions in [^{15}N, ^{1}H]- and [^{13}C, ^{1}H]-HSQC spectra. Side-chain resonance assignment of TM67 was accomplished using the following spectra: hCCH-TOCSY[29,30] in combination with [^{13}C,^{1}H]-HSQC and ^{13}C-resolved aliphatic-NOESY[31] experiments. The start of the side-chain assignment is by using the known Cα and Cβ chemical shifts obtained from the backbone assignments.

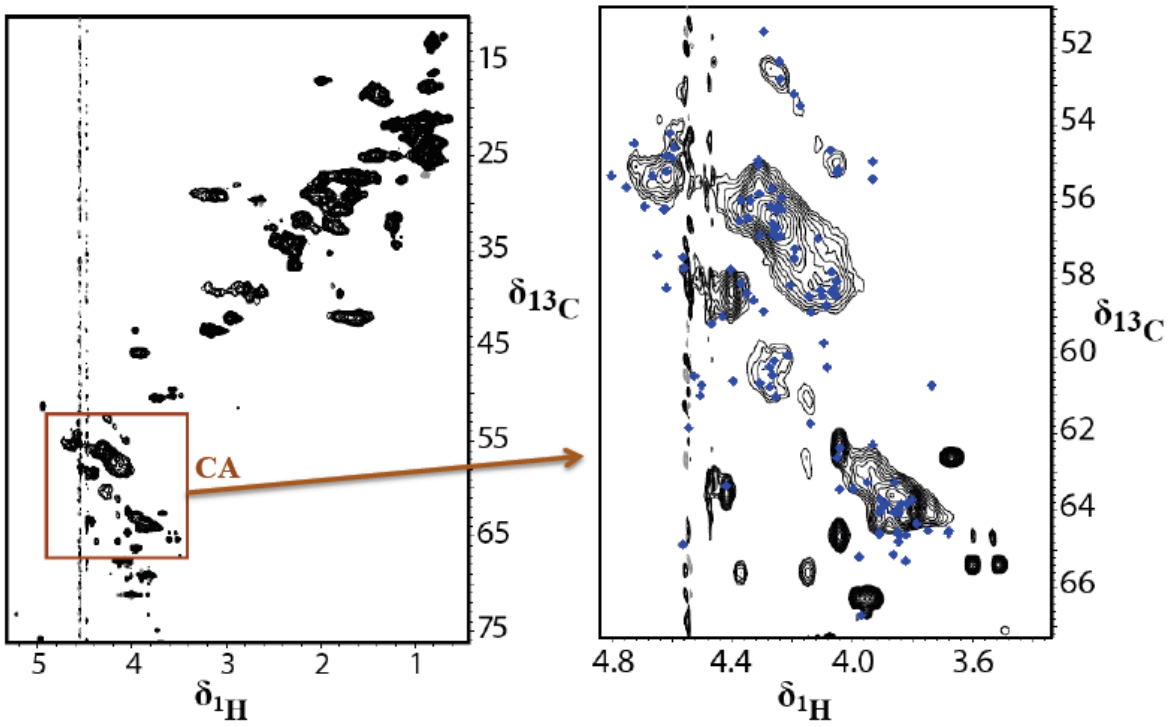


Figure 4: [^{13}C,^{1}H]-HSQC spectrum of TM67. The area in the box corresponds to Cα chemical shifts and on enlarging (right) we see significant overlap of peaks. The blue crosses indicate the Cα peak positions from the backbone assignment. These assignments are initially made using the protein labeled with ^{15}N, ^{13}C and ^{2}H. When transferred onto [^{13}C,^{1}H]-HSQC spectrum, some of the labels are distant to the resonances denoting a shift in peak positions with the new NMR sample or when labeled differently. Alternatively, the peaks may be attenuated by very efficient T2 relaxation.

In CARA, by using ^{13}C-hCCH-TOCSY spectrum in systemscope [rotate: D1-Y(^{1}H):D2-X(^{13}C):D3-Z(^{13}C)], we obtain the C-C-H chemical shifts, i.e. Hα and Hβ from Cα and Cβ respectively. This is followed by rotating to D1-Xaxis(H):D2-Zaxis(C):D3-Yaxis(C) for obtaining the H-C-C chemical shifts, i.e. Cγ, Cδ and Cε. Once the C-spin system is

assigned, using the procedure applied to obtain Hα and Hβ, *protons of* γ, δ, and ε are assigned. Assigned protons were further confirmed using the ^{15}N-resolved NOESY data. More detailed explanation on side-chain assignment is explained exclusively on the CARA wiki page. Although, I have extensively used the ^{13}C-resolved and ^{15}N-resolved NOESY spectra data to overcome heavy overlap of peaks from the same type of aminoacid, we were not able to assign few of the isolated peaks observed in the [^{13}C, ^{1}H]-HSQC spectra (figure 4). Probably, these unassigned peaks could be attributed to the unidentified residues in the backbone assignment. Due to high overlap of the peaks as observed from the [^{13}C, ^{1}H]-HSQC spectra, I have not attempted to progress assigning the proton chemical shifts of $-CONH_2$ amides corresponding to Asn and Gln and the aromatic side-chains of Tyr, Trp and Phe. At this stage, I stopped working on the side-chain assignment of TM67 and progressed onto different fragments.

3.4 Construct 2: His-ECL2-TM56-ECL3 (HisTM56):

The main idea behind designing this construct was to locate the resonances of unassigned residues of TM6 in TM67. Fortunately, HisTM56 construct could be directly expressed without the requirement of a fusion tag. Expression and purification procedures are explained in detail in the previous chapter. NMR sample conditions were following the procedures for optimization as explained for TM67. HisTM56 protein was solubilized in 40 mM phosphate buffer consisting of 5% LPPG and 40 mM DTT. For backbone assignments, a uniformly labeled ^{15}N, ^{13}C and ^{2}H protein was produced and purified. A [^{15}N, ^{1}H]-TROSY spectrum and triple-resonance spectra such as HNCO, HNCACO, HNCA, HNCOCA, HNCACB and HNCOCACB were recorded for backbone assignments. NMR acquisition parameters for each experiment were reported in a table 2. Spectra were processed within the Bruker Topspin software and converted to XEASY format to perform chemical shift assignment using the software CARA (Keller 2004).

HisTM56								
	Data points			Spectral width			NS	
	F3	F2	F1	F3	F2	F1		

[^{15}N,^{1}H]-HSQC		2048	256		18	22	32	
HNCO	2048	40	110	18	22	20	8	
HN(CA)CO	2048	40	110	16	22	20	16	
HNCACB	2048	40	128	15	22	70	16	
HN(CO)CACB	2048	40	128	15	22	70	16	
HNCA	2048	40	128	16	22	30	16	
HN(CO)CA	2048	40	128	16	22	30	16	
^{15}N-NOESY	2048	40	160	14	22	10	32	100ms mix

Table 2: Spectroscopic details of NMR experiments used for measuring HisTM56

Unfortunately, only 72 out of the expected 132 could be assigned (figure 5 & 6). Originally, HisTM56 construct comprises of 132 residues starting with a Met followed by an 8x His-tag and includes 4 Pro's. We again attribute the missing assignments to either line broadening or peak overlap but unlike the previous construct we still find strong peaks that are not assigned.

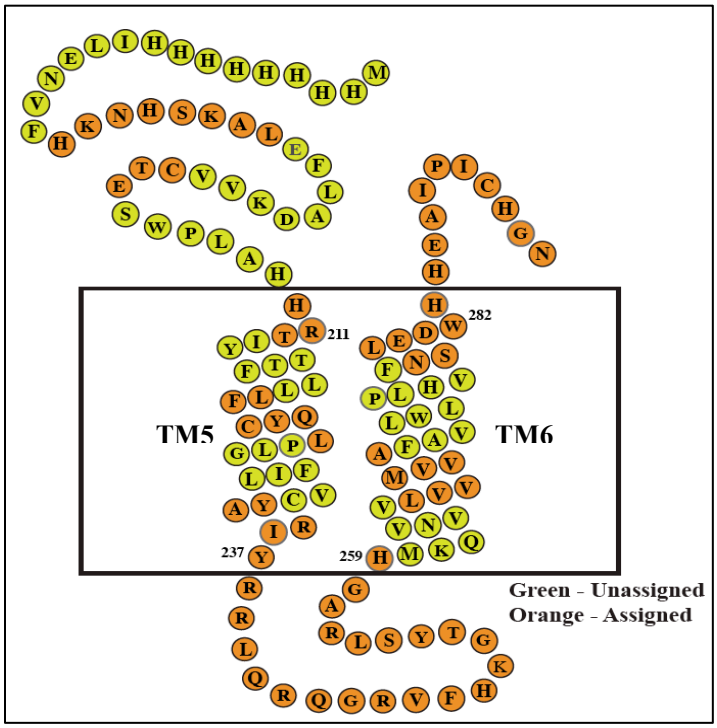


Figure 5: Snakeplot of HisTM56 with the assigned residues in orange and the unassigned in green. In total 72 residues were assigned out of the 132 residues present in HisTM56 construct.

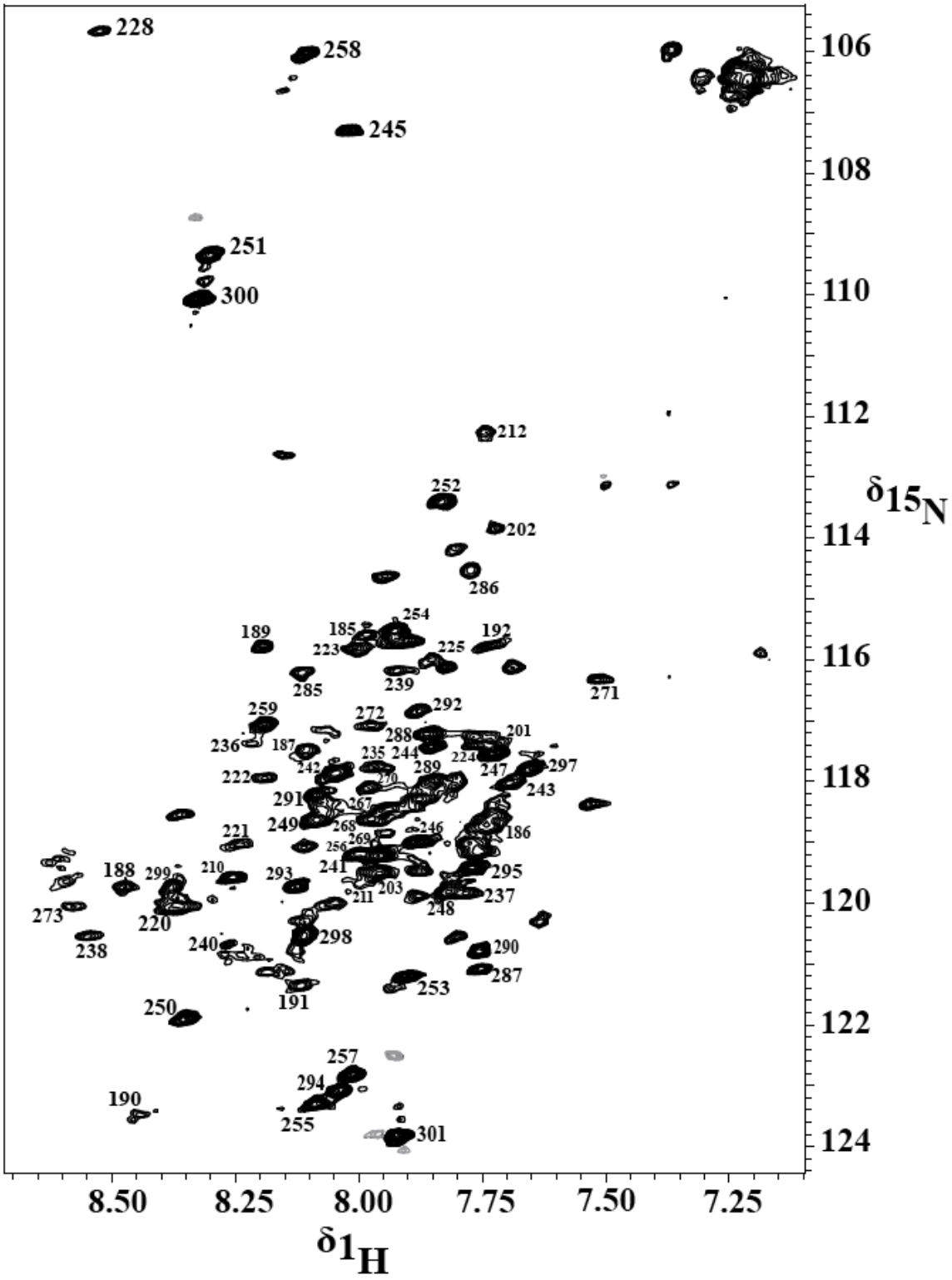


Figure 6: [^{15}N,^{1}H]-TROSY spectrum of HisTM56 with the assigned resonances. Numbers indicated on top of each resonance represents the corresponding aminoacid residue present in the full-length NPY4R.

The signal to noise ratio for a few triple resonance experiments was very low due to their measurements with lesser number of scans, causing difficulty in peak-picking close to the strong resonances observed in the [^{15}N, ^{1}H]-TROSY spectrum. This also resulted in identifying the succeeding or preceding residue matches difficult. A noteworthy observation in HisTM56 assignment was that the HNCACO spectrum allowed picking only half of the signals because of poor S/N. Thus we eliminated matching of carbonyl chemical shifts and relied only on Cα and Cβ chemical shifts. Although protein concentration was sufficient (0.4 mM), we believed that the presence of the His-tag on the N-terminus resulted in conformational exchange, possibly causing line-broadening effects leading to a decrease in signal to noise.

Matching peak-picked resonances in TM67 for assignment was pretty straightforward unlike HisTM56. Difference between the two constructs apart from the NPY4R aminoacid sequence is the presence of a His-tag in TM56. Probably, this was causing conformational changes and thus line-broadening resulting in missing of few resonances in HisTM56. To verify the effect of the N-terminal His-tag, we designed a new construct with the same aminoacid sequence but with a His-tag at the C-terminus. This new construct was expressed in very low quantities in light water and gave absolutely no expression in deuterated water. Because of the limitations in expression of TM56His, we were only able to compare the [^{15}N,^{1}H]-TROSY spectrum with that of HisTM56 (figure 7). On comparison, we observed a significant shift changes for few peaks, but most resonances remained the same. We noticed that TM56His resulted in better signal to noise ratio with more homogeneous line-widths compared to the prior construct HisTM56. The [^{15}N,^{1}H]-TROSY of TM56His displayed 105 out of the expected 130 peaks in total. We were not able to go any further because of the lack of expression in deuterated water.

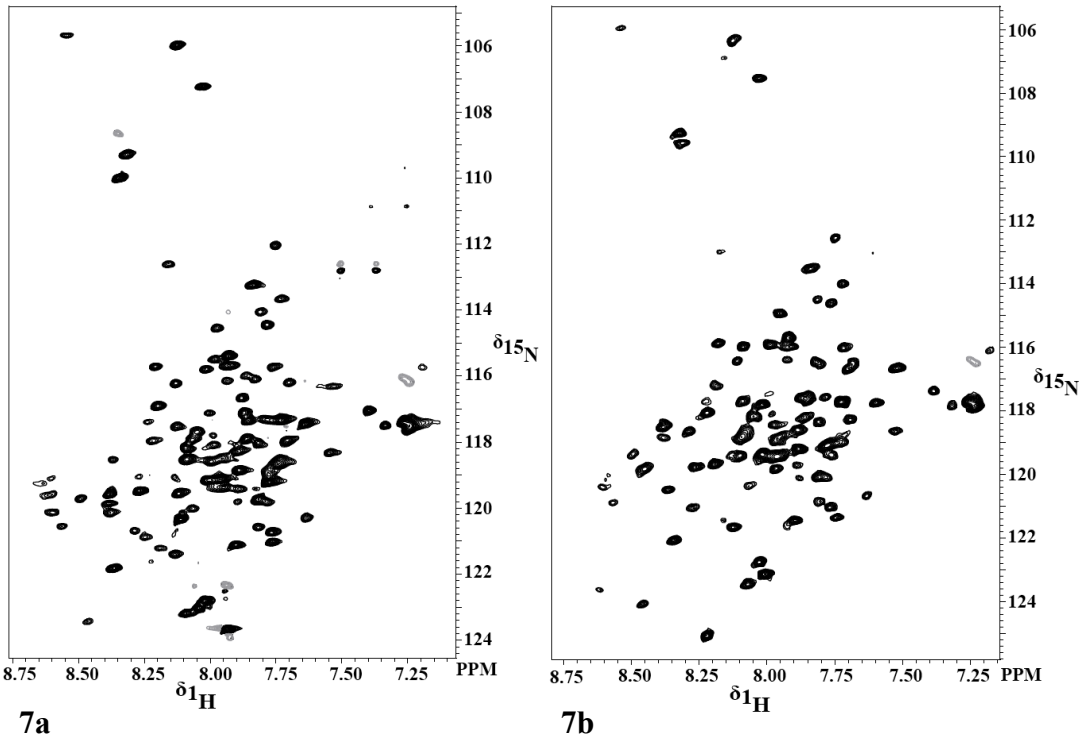


Figure 7: [^{15}N,^{1}H]-TROSY spectra of HisTM56 (left) and TM56His (right) solubilized in 5% LPPG.

3.5 Construct 3: ECL2-TM567-Cter-His without Cysteines (Cys-free TM567His):

Before working on the Cys-free mutant of TM567, I have expressed, purified and analyzed the wild-type TM567His by NMR. Expression of this larger construct was possible directly without the need for a fusion tag. Wild-type TM567His had 4 Cys residues, which were later mutated to avoid aggregation due to disulfide bond formation. Significant yields of purified protein were obtained for both the wild-type and the Cys-free variants of TM567 in light water, while only for the Cys-free form sufficient quantities could be obtained from the cultures grown in deuterated water. A 0.34 mM NMR sample (^{15}N, ^{13}C and ^{2}H labeled) of wild-type was prepared and an attempt was made to analyze the assignment using the triple-resonance spectra.

Some of the NMR experiments such as the [^{15}N,^{1}H]-TROSY, HNCO, HNCA and HNCOCA resulted in good spectra, other experiments such as HNCACO, HNCOCACB and HNCACB contained only a subset of the expected peaks. Initially, I tried to match

resonances using just the HNCA and HNCOCA but the strip matching was too complicated since too many ambiguities remained. A little progress in assignment (30 out of 206 residues) was made with the wild-type TM567His construct. At the same time, the Cys-free mutant of TM567His was designed to facilitate recording of PREs. The Cys mutant could be obtained in sufficient quantities in deuterated water, and an NMR sample was prepared using the same procedures applied to the previous constructs. A [^{15}N,^{1}H]-TROSY measurement on the mutant sample displayed peaks with better signal to noise and more homogenous line-widths when compared to the spectrum of wt-TM567(figure 8).

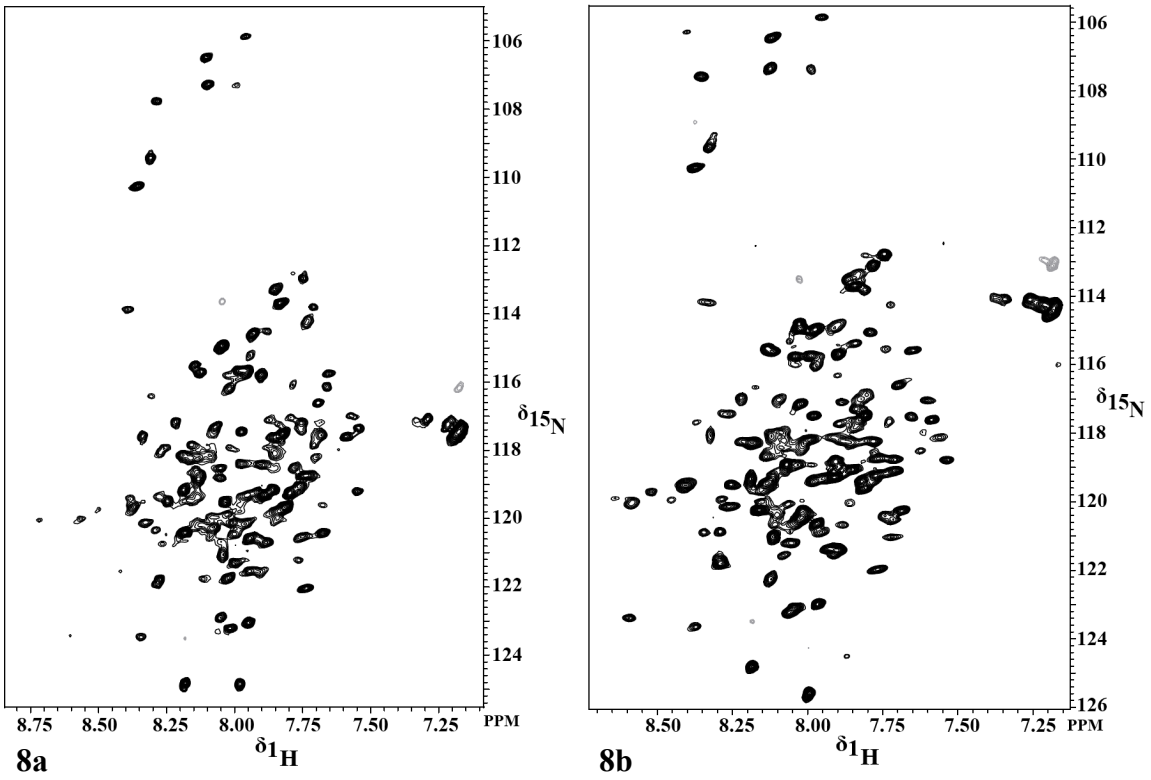


Figure 8: [^{15}N,^{1}H]-TROSY spectra of wt-TM567His (left) and Cys-free TM567His (right) solubilized in 5% LPPG.

For the backbone assignment a [^{15}N, ^{1}H]-TROSY spectrum and triple-resonance spectra such as HNCO, HNCACO, HNCA, HNCOCA, HNCACB and HNCOCACB were recorded. NMR acquisition parameters for each experiment were reported in a table 3.

Spectra were processed within the Bruker Topspin software and converted to XEASY format to perform chemical shift assignment using the software CARA (Keller 2004).

Cys-free TM567His								
	Data points			Spectral width			NS	
	F3	F2	F1	F3	F2	F1		
[^{15}N,^{1}H]-HSQC		2048	256		18	30	32	
HNCO	2048	40	100	18	22	16	16	
HNCACO	2048	40	90	16	22	18	32	
HNCACB	2048	40	128	16	22	69	32	
HN(CO)CACB	2048	40	140	18	22	69	16	
HNCA	2048	40	90	18	22	30	32	
HN(CO)CA	2048	40	90	18	22	30	32	
^{15}N-NOESY	2048	44	140	14	22	10	24	100ms mix

Table 3: Spectroscopic details of NMR experiments used for measuring Cys-free TM567His.

A procedure for backbone assignments was followed as described for TM67. All triple resonance spectra gave reasonable correlations. Around 120 aminoacids were assigned out of a total of 206 in the Cys-free TM567His construct (figure 9 & 10). Most assignments stem from resonances from the soluble loops and the termini. Almost the complete N- and C-termini have been assigned. A significant region in TM7 and a small stretch in TM6 are also assigned. None of the residues from the TM5 have been assigned. Apart from the assigned resonances, there exist many unassigned strong peaks in the [^{15}N,^{1}H]-TROSY spectrum (figure 9). These unassigned strong resonances show correlations in triple resonance spectra, but could not be assigned because no confident succeeding or preceding matches were found.

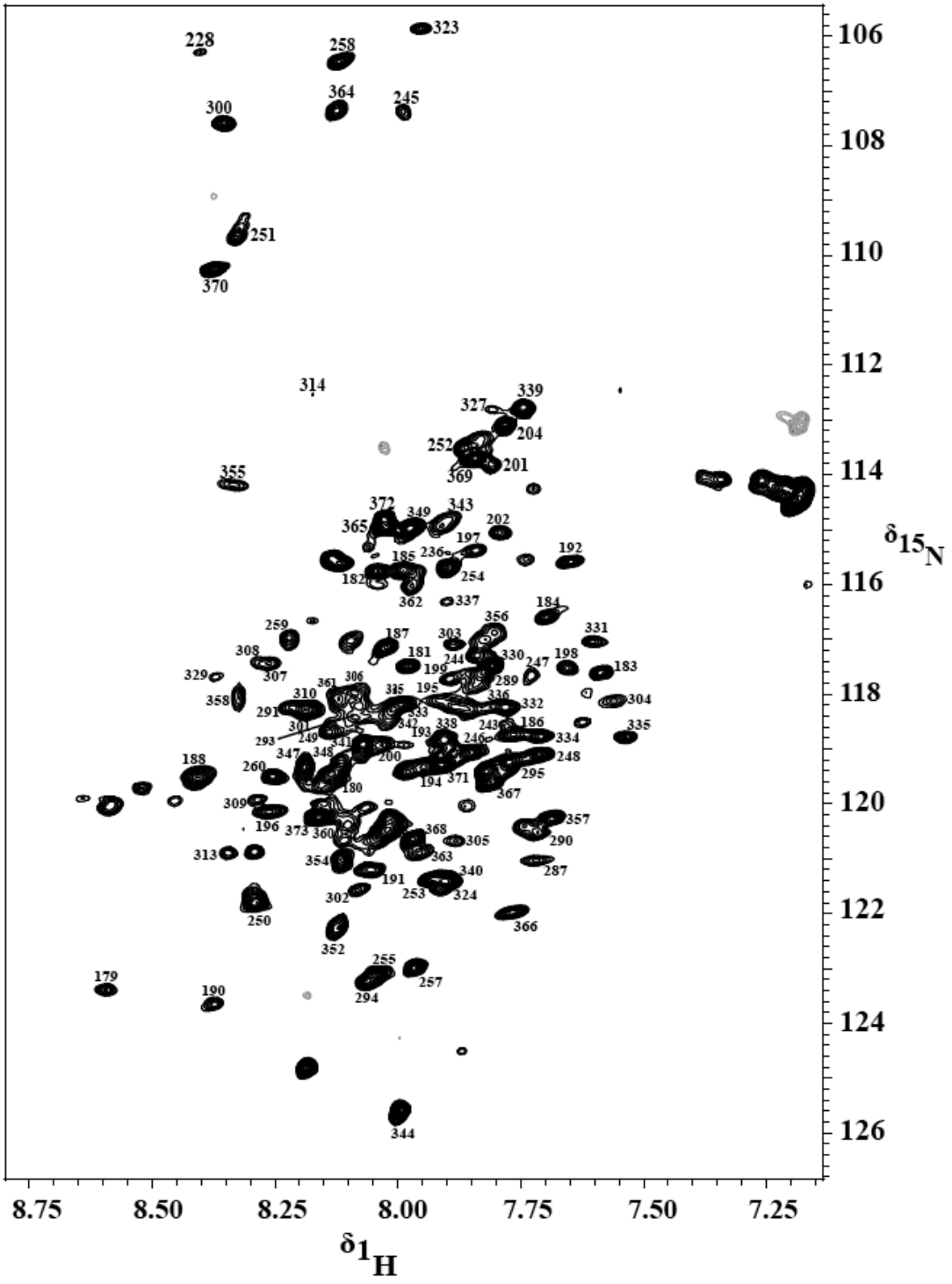


Figure 9: [^{15}N,^{1}H]-TROSY spectrum of Cys-free TM567His with annotated assignments. Numbers indicated on top of each resonance correspond to numbering in full-length NPY4R.

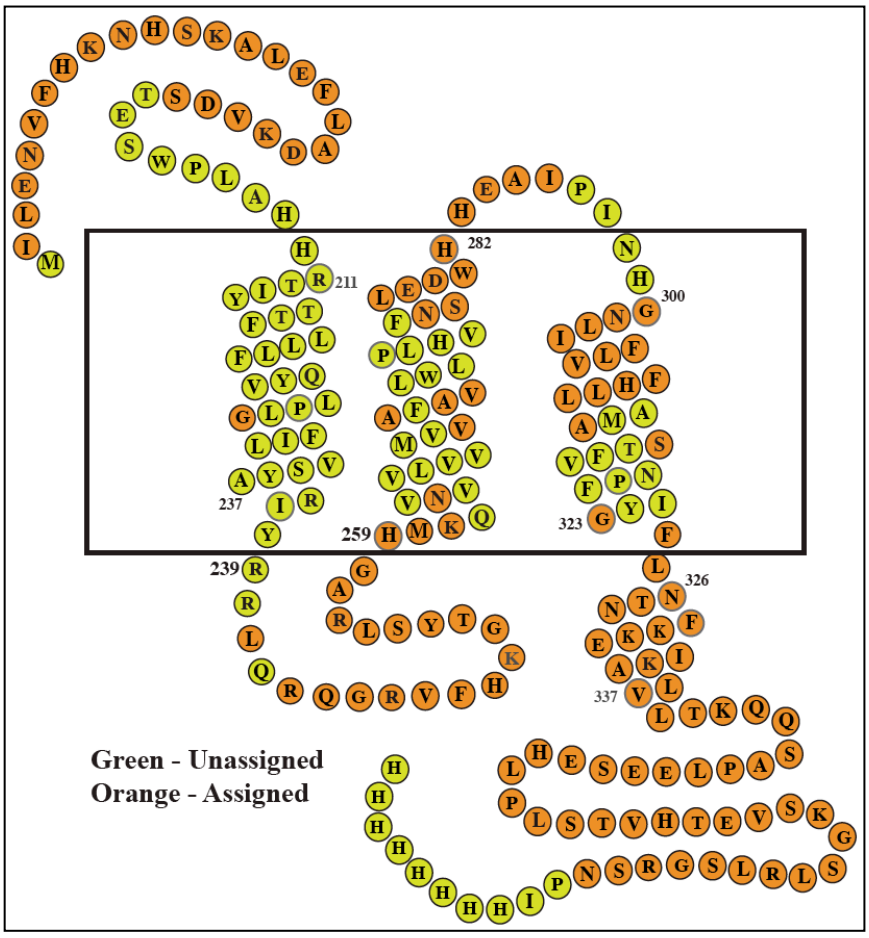


Figure 10: Snakeplot of Cys-free TM567His with the assigned residues in orange and the unassigned in green. In total 110 residues were assigned out of the 206 residues present in the construct.

Finding matches to the unassigned peaks was complicated both for the remaining strong and weak peaks. In a few instances, matches for these peaks were found with the already assigned peaks. In few cases, isolated strong resonances were assigned by comparing their ^{15}N, ^{1}H chemical shifts and ^{15}N-NOESY correlations with the assigned peaks in spectra's of TM67 and HisTM56. One another strategy where assignments were identified to the unassigned strong peaks was by finding its succeeding or preceding partner on matching all strips in stripscope and identifying the labels that were either not picked or picked wrongly.

The central region of [^{15}N,^{1}H]-spectrum shows poor signal dispersion and the peak picking process in this region was performed based on the HNCO correlations. The HNCO experiment is very sensitive and sometimes we found more peaks than observed in the [^{15}N,^{1}H]-TROSY spectrum or recognized peak overlap. These extra peaks identified in the HNCO correlations don't show any connections in the other triple-resonance spectra.

Similar to the previous cases, most of the missing assignments were due to absence of resonances as an effect of line-broadening or because of peak overlap. Peak overlap in the [^{15}N,^{1}H]- spectrum could in principle be overcome by observing the correlations obtained in HNCO or in any other triple-resonance spectrum, but the problem still persists for repetitive sequences of the same aminoacid type. This problem exists for example in TM6, which comprises repetitive Val stretches. In some cases, identifying Gly or Ala based on their unique chemical shifts provided anchor points for short additional stretches of assignments.

All permutations and combinations were tried for assigning as many residues as possible with the use of ^{15}N-NOESY as a verification tool.

4. Discussion

Apart from the quantities of TMPs required and the sensitivity of NMR technique, other factors that are critical for the biophysical analysis of TMPs by NMR are (1) the sample preparation and (2) signal overlap due to repetitive aminoacid residues present in the TM region and (3) absence of signals due to signal broadening due to the size of the protein-detergent complex or due to conformational exchange. The former is due to the large size of a complex of polytopic TMP within detergents or lipids, together with the inherent mobility of TMP's helical bundle that causes fast relaxation and strong non-uniform broadening of NMR signals and, as a result, several problems with the assignment, spectra analysis, and identification of long-range interactions. The weight limitations can only be bypassed with the invention of new methods suitable for such large entities. Some recent technological developments in cell-free protein synthesis, specific selective

isotope labeling, systematic paramagnetic labeling and data analysis, have already made the elucidation of many polytopic α-helical TMP structures possible[2,3,32], but many more are necessary. Generally, it is the tendency of α-helical TMPs to result in overlapping or crowded spectra due to the presence of repetitive hydrophobic aminoacids, which have analogous chemical shifts. Maybe in future development of high-field magnets could help in signal dispersion solving crowded spectra problem.

Problems and their troubleshooting associated with membrane protein NMR can be interpreted using 3 major considerations:

1. **Labeling strategies:** Selective labeling, segmental labeling and high deuteration
2. **Membrane mimetics:** Conformational exchange processes, generating stabilized receptor mutants or optimizing detergents/lipids for suitable membrane environment.
3. **NMR experiments:** Due to large size of TMPs, shorter T2's don't display necessary correlations in triple resonance spectra

4.1 Labeling Strategies: To achieve almost complete backbone assignments of the fragments despite the presence of substantial signal overlap, aminoacid type selective labeling schemes might be helpful, but so far we did not do so because of sub-optimal yields of our fragments in deuterated water. In future, we aim to optimize the yields by taking advantage of cell-free expression system, which could help in assigning the missing residues because it allows selective labeling on amino acids. Selective labeling in cell-free expression is considerably cheaper allowing labeling of residues - AFGILV, which predominantly occur in the TM regions. ^{15}N and ^{13}C labeling strategies of the three specific aminoacids (ILV) primarily found in the transmembrane parts of TMPs can be achieved by using aminoacid precursors[33]. Although segmental labeling techniques are popular in case of soluble proteins and ideal for our divide and conquer approach, it is hardly implemented for TMPs. Other strategies for the labeling of TMPs have been developed for specific NMR experiments that help in backbone assignments[34].

4.2 Membrane mimetics and TMP sample preparation for NMR: I believe NMR sample preparation should be a standardized process as preparing samples with

differently labeled proteins always resulted in chemical shift changes. This needs chemical shift adaptations, which is a tedious process. Sometimes also inhomogeneous line-widths were observed, especially in the case of doubly labeled TM67 compared to the triply labeled sample. This might also be a result of different protein-detergent interactions, which is deuterated in the sample preparation of doubly labeled protein and protonated in the later case. I note here that deuterated compounds are slightly more hydrophobic when compared with the protonated counterparts. The result of inhomogeneity in line widths could be attributed to the conformational changes as a result of protein-detergent interactions within the micelle. This affects the overall protein stability, and the conformational exchange involving a few residues or regions leading to exchange broadening. Also the protein interactions with the deuterated detergent molecules could give rise to variations in TMP packing leading to chemical shift changes.

4.3 NMR Experiments: The sensitivity of NMR experiments used for assignment varies a lot with few being very sensitive and much more insensitive. This puts burden on the biochemists demanding more quantities of TMPs for their analysis. In the previous chapter, I discussed the challenges in obtaining purified TMPs and also the effects in protein yields, which drastically vary when changing medium from light to heavy water. In all the triple-resonance experiments, HNCO is the most sensitive experiment whereas its counter experiment HN(CA)CO is only $1/10^{th}$ in sensitivity to HNCO[35]. Thus in most cases we faced difficulties when using the peaks picked from HN(CA)CO for strip matching because many correlations are missing. In my experience the HN(CO)CACB is more sensitive than the HNCACB and there were issues with the Cα and Cβ peak picking from HNCACB spectra because few of the resonances were at the border of noise to pick them unambiguously. Lack of correlations from Cα were compensated by measuring the more sensitive HN(CO)CA and HNCA experiments. We understand that the sensitivity component could be compensated on using higher concentrations of proteins, which is a limitation for proteins expressed in deuterated water. An alternative is to increase the number of scans, which sometimes doesn't show any improvement in signal to noise ratio. With 16 or 32 scans measurement time for each triple resonance experiment is around 1.5 or 3 days. Thus, significantly improving S/N by increasing the number of scans significantly is not really possible. From my experience, it is only worth to measure

triple-resonance experiment with protein concentrations is at least 0.4 mM. Any concentration less than the mentioned would be a significant waste of resources.

To conclude, we imagine that the studies on fragments of GPCRs containing TM domains may be beneficial because they support investigation of the recruiting and folding processes in the membranes, as explained in the two-stage model established by Popot and Engelman[36,37]. We believe that the C-terminal fragments of NPY4R do integrate into the detergent micelles, but their active form and stable state could only be obtained in presence of their partnering helices. Possibility that the residues were unable to be assigned by NMR could be attributed to the instability of shorter fragments with the need to search for their stabilizing partners (helices), which is resulting in conformational exchanges and thus line broadening, causing disappearance of resonances.

From my experience, biochemistry of purifying various NPY4R fragments or TMPs was though challenging, it could definitely be overcome if the protein shows expression in the *E.coli* system. Biophysical analysis of the purified fragments or TMPs by NMR needs further optimization either in selecting better membrane mimetic or by increasing the sensitivity of NMR experiments. Protein expression, membrane mimetic optimization and NMR techniques improvement are all at the moment transforming or advancing and hope in the near future these will no more pose a challenge for structural biologists.

5. References

1. Park, S. H. et al. Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* **491**, 779-783 (2012).
2. Reckel, S. et al. Solution NMR structure of proteorhodopsin. *Angew Chem Int Ed Engl* **50**, 11942-11946 (2011).
3. Berardi, M. J., Shih, W. M., Harrison, S. C. & Chou, J. J. Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching. *Nature* **476**, 109-113 (2011).
4. Gautier, A., Mott, H. R., Bostock, M. J., Kirkpatrick, J. P. & Nietlispach, D. Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nat Struct Mol Biol* **17**, 768-774 (2010).
5. Bayrhuber, M. et al. Structure of the human voltage-dependent anion channel. *Proc Natl Acad Sci U S A* **105**, 15370-15375 (2008).
6. Van Horn, W. D. et al. Solution nuclear magnetic resonance structure of membrane-integral diacylglycerol kinase. *Science* **324**, 1726-1729 (2009).

7. Hiller, S. et al. Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. *Science* **321**, 1206-1210 (2008).
8. Zhou, Y. et al. NMR solution structure of the integral membrane enzyme DsbB: functional insights into DsbB-catalyzed disulfide bond formation. *Mol Cell* **31**, 896-908 (2008).
9. Baker, K. A., Tzitzilonis, C., Kwiatkowski, W., Choe, S. & Riek, R. Conformational dynamics of the KcsA potassium channel governs gating properties. *Nat Struct Mol Biol* **14**, 1089-1095 (2007).
10. Sanders, C. R. & Landis, G. C. Reconstitution of membrane proteins into lipid-rich bilayered mixed micelles for NMR studies. *Biochemistry* **34**, 4030-4040 (1995).
11. Raschle, T. et al. Structural and functional characterization of the integral membrane protein VDAC-1 in lipid bilayer nanodiscs. *J Am Chem Soc* **131**, 17777-17779 (2009).
12. Kielec, J. M., Valentine, K. G., Babu, C. R. & Wand, A. J. Reverse micelles in integral membrane protein structural biology by solution NMR spectroscopy. *Structure* **17**, 345-351 (2009).
13. Bekei, B. et al. In-cell NMR in mammalian cells: part 3. *Methods Mol Biol* **895**, 67-83 (2012).
14. Selenko, P. & Wagner, G. Looking into live cells with in-cell NMR spectroscopy. *J Struct Biol* **158**, 244-253 (2007).
15. Homans, S. W. NMR spectroscopy tools for structure-aided drug design. *Angew Chem Int Ed Engl* **43**, 290-300 (2004).
16. Vogtherr, M. & Fiebig, K. NMR-based screening methods for lead discovery. *EXS* 183-202 (2003).
17. Tugarinov, V., Hwang, P. M. & Kay, L. E. Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annu Rev Biochem* **73**, 107-146 (2004).
18. Sanders, C. R. & Sonnichsen, F. Solution NMR of membrane proteins: practice and challenges. *Magn Reson Chem* **44 Spec No**, S24-S40 (2006).
19. Ratnala, V. R. New tools for G-protein coupled receptor (GPCR) drug discovery: combination of baculoviral expression system and solid state NMR. *Biotechnol Lett* **28**, 767-778 (2006).
20. Nielsen, N., Malmendal, A. & Vosegaard, T. Techniques and applications of NMR to membrane proteins. *Mol Membr Biol* **21**, 129-141 (2004).
21. Lindert, S. et al. Drug screening strategy for human membrane proteins: From NMR protein backbone structure to in silica- and NMR-screened hits. *Biochem Biophys Res Commun* **445**, 724-733 (2014).
22. Horst, R., Liu, J. J., Stevens, R. C. & Wuthrich, K. beta -Adrenergic Receptor Activation by Agonists Studied with F NMR Spectroscopy. *Angew Chem Int Ed Engl* (2013).
23. Cherezov, V. Lipidic cubic phase technologies for membrane protein structural studies. *Curr Opin Struct Biol* **21**, 559-566 (2011).
24. Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P. & Landau, E. M. X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science* **277**, 1676-1681 (1997).
25. Landau, E. M. & Rosenbusch, J. P. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc Natl Acad Sci U S A* **93**, 14532-14535

(1996).
26. Kocherla, H. et al. Biosynthesis and spectroscopic characterization of 2-TM fragments encompassing the sequence of a human GPCR, the Y4 receptor. *Chembiochem* **13**, 818-828 (2012).
27. Zou, C., Naider, F. & Zerbe, O. Biosynthesis and NMR-studies of a double transmembrane domain from the Y4 receptor, a human GPCR. *J Biomol NMR* **42**, 257-269 (2008).
28. Shao, X., Zou, C., Naider, F. & Zerbe, O. Comparison of fragments comprising the first two helices of the human Y4 and the yeast Ste2p G-protein-coupled receptors. *Biophys J* **103**, 817-826 (2012).
29. Ikura, M., Kay, L. E. & Bax, A. Improved three-dimensional 1H-13C-1H correlation spectroscopy of a 13C-labeled protein using constant-time evolution. *J Biomol NMR* **1**, 299-304 (1991).
30. Olejniczak, E. T., Xu, R. X. & Fesik, S. W. A 4D HCCH-TOCSY experiment for assigning the side chain 1H and 13C resonances of proteins. *J Biomol NMR* **2**, 655-659 (1992).
31. Michael Sattler, Jurgen Schleucher & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog. Nucl. Magn. Reson. Spectrosc. **34**, 93-158 (1999).
32. Klammt, C. et al. Facile backbone structure determination of human membrane proteins by NMR spectroscopy. *Nat Methods* **9**, 834-839 (2012).
33. Tugarinov, V. & Kay, L. E. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc* **125**, 13868-13878 (2003).
34. Reckel, S. et al. Transmembrane segment enhanced labeling as a tool for the backbone assignment of alpha-helical membrane proteins. *Proc Natl Acad Sci U S A* **105**, 8262-8267 (2008).
35. Lescop, E., Schanda, P. & Brutscher, B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J Magn Reson* **187**, 163-169 (2007).
36. Popot, J. L. & Engelman, D. M. Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem* **69**, 881-922 (2000).
37. Popot, J. L. & Engelman, D. M. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **29**, 4031-4037 (1990).

Chapter V: Incorporation of TM567His Fragment of NPY4R into Nanodiscs

1. Abstract

Generally, Transmembrane Proteins (TMPs) perform their biological functions being embedded in their natural membrane environment, which are phospholipid bilayers that in addition contain other proteins as well as glycolipids. The natural phospholipid bilayers are highly dynamic and heterogeneous in composition, limiting the study of TMPs by standard structural and biophysical techniques such as X-ray crystallography, NMR, EM and CD. To provide a chemically more defined environment, TMPs are extracted into detergents or lipids *in vitro*. In these artificial environments, TMPs face various challenges rendering them being functionally inactive, or existing in aggregated form, and in addition they may be less stably folded or and solubility issues might be introduced. Nevertheless, some of these environments can closely mimic properties of native environments. One such environment recently developed by Sligar and colleagues is very promising, and is the incorporation of TMPs into nanodiscs. Nanodiscs are made up of discoidal phospholipid bilayers (~ 10-12 nm large) that are stabilized by the packing of amphipathic helical membrane scaffold protein (MSP) around them. A TMP incorporated into nanodiscs is claimed to remain stable, soluble and mono-disperse with the ability to access both sides of the lipid bilayer. These properties render scientists to perform structural, biophysical and biochemical investigations in a more effective and simpler way.

On the contrary, although nanodiscs mimic near-native environments, their large particle size complicates the spectral analysis of TMPs by solution NMR. Recently this disadvantage was tried to overcome by developing smaller nanodiscs, which are obtained by truncating the MSP. However, it needs to be demonstrated yet that these are well suited for multi-spanning TMPs. In this chapter, we describe the methodology for incorporating the 3-TM fragment TM567 of NPY4R into nanodiscs of smaller diameter.

* [This project was done in collaboration with my fellow colleague Mr. Philipp Ansorge. PhD student in the same lab]

2. Introduction

To completely understand the biology of a TMP, biochemical conclusions relating to their functional role (e.g. signal transduction) need to be substantiated by structural studies (e.g. NMR or X-ray studies). Although tremendous progress in understanding the function of TMPs has been achieved, limitations still exist due to the difficulties associated with obtaining sufficient yields and incorporating them into a membrane mimetic environment in a functional form. To choose the best membrane mimetic, first we should understand the nature of native cell membranes in which the TMPs naturally exist. Host cell membranes are made up of phospholipid bilayers in which lipid molecules not only play a role in forming physiological membranes enveloping cell contents and embedding TMPs but also act in modulating conformational rearrangements of these proteins[1-4]. For this reason, biochemical studies of TMPs *in vitro* should be carried out with lipid molecules that closely resemble phospholipids in bilayers of cell membranes.

Phospholipids constitute the major components in cell membranes along with many other minor components (cholesterol, proteins and glycolipids). They are amphipathic with a hydrophilic head group (negatively charged phosphate) and a hydrophobic tail (long hydrocarbon chain). These lipids when placed in water, line up against each other and also on top of each other in antiparallel fashion, resulting in bilayers with the hydrophilic head groups exposed to water and the hydrophobic tails buried inside the bilayer (fig 1).

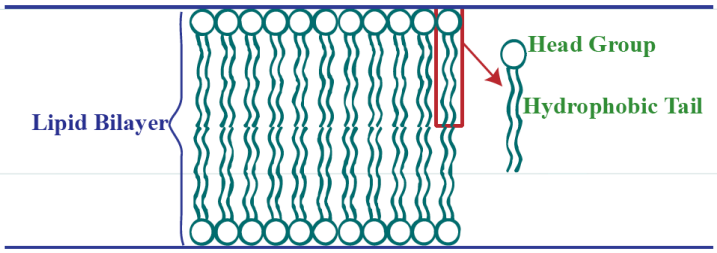


Figure 1. Schematic representation of a lipid bilayer indicating the arrangement of phospholipid molecules. Each lipid molecule is made up of a hydrophilic head group and a hydrophobic tail.

Natural lipid bilayers in cell membranes are considered to be extremely porous or dynamic allowing the proteins to constantly move and interact with other species present in the same bilayer. These movements within lipidic membranes can lead to aggregation and sometimes give rise to false biological outcomes *in vitro*[5]. Also, a biophysical and structural study on TMPs present in dynamic lipid bilayers or liposomes is extremely difficult. To minimize these effects one has to come up with a design of membranes models identical with lipid bilayers, but which are more rigid and which keep the target TMP in a monodisperse form. Although many artificial membranes such as micelles, bicelles and artificial lipid bilayers (liposomes) could satisfy the property of native-like membrane environment, none of these satisfy the property of rigidity or inflexibility and monodispersity[6,7] or the requirement in terms of maximum size of the system.

A new technology has been developed that relies on reconstituting TMPs in discoidal-shaped nanometer-sized soluble phosopholid bilayer membranes termed as nanodiscs or nanolipoprotein particles (NLPs). These assemblies mostly satisfy the ideal properties mentioned above for studying the TMPs *in vitro* and for structural analysis. Incorporating TMPs into nanodiscs was first pioneered by Sligar and colleagues by taking advantage of the amphipathic properties of apolipoprotein A-1 (ApoA-1), which acts as a belt/scaffold surrounding and stabilizing the lipid bilayer[8,9]. Thus ApoA-1 is also referred to as the membrane scaffolding protein (MSP)[10,11] (fig 2). Based on the amphipathic property of MSPs, its hydrophobic surface interacts with the long fatty acyl side-chains of the lipid molecules within the bilayer, and the polar surface interacts with the aqueous environment making the whole nanodisc entity soluble. Formation of a nanodisc with or without an incorporation of a TMP is a spontaneous self-assembly process regulated by the nature of the MSP, and the speed at which detergent is removed. Once nanodiscs are formed, they are highly stable entities in solution in contrast to the highly dynamic micelles formed by detergents. Thus incorporating TMPs into nanodiscs adds tremendous advantages such as more closely mimicking the natural lipid bilayer environment, and the fact that they possibly incorporate proteins in monodisperse form.

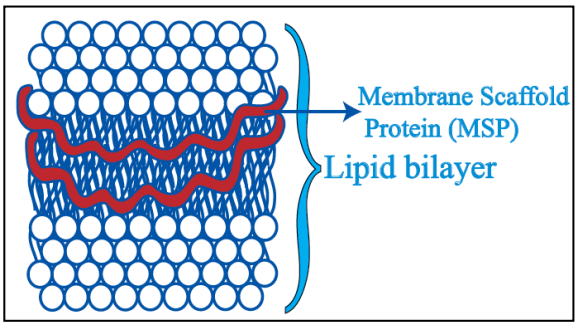


Figure 2: Model of an empty Nanodisc. Basic constituents comprise the lipid molecules that form the bilayer, surrounded by 2 molecules of membrane scaffold protein.

The size of the nanodisc can be adjusted by using different MSP variants to either accommodate either large multi-spanning TMPs or TMPs with very few multi-spanning domains[12]. The most commonly used MSP, ApoA-1, is made up of 10 α-helices. For generating smaller nanodiscs, some of the helices can be removed to generate shorter version MSPs resulting in smaller nanodiscs (fig 3). The shortest version of MSP produces discs of approximately 10 nm in diameter with a size estimated to be 150-200kDa.

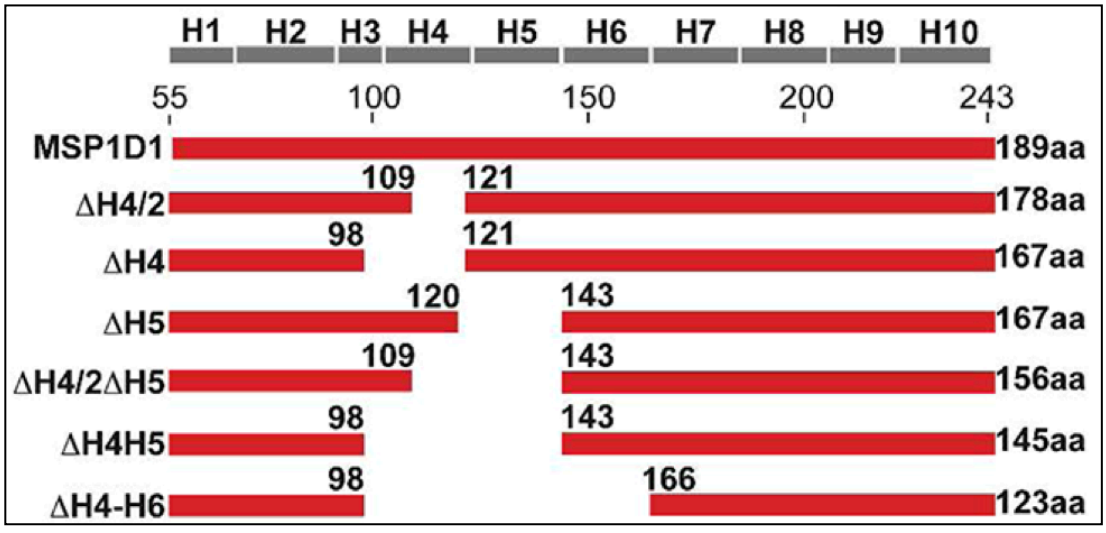


Figure 3. Different versions of MSPs used in generating nanodiscs of different sizes[13].

Today, this technique has been validated for performing biochemical experiments *in vitro,* and is increasingly used to incorporate a variety of TMPs. Some of the TMPs that

were successfully incorporated into nanodiscs include VDAC-1[14], VDAC-2[15], CD4mut[16], the voltage-sensing domain of the potassium channel KvAP[17], rhodopsin[18], bacteriorhodopsin[19] and cytochrome P450[9]. In this chapter, we describe the process of incorporating a Cys-free mutant of the 3-helical transmembrane fragment TM567 of NPY4R, which is a G-protein Coupled Receptor (fig 4a and 4b). We highlight the importance of selecting the best ratios of MSP to lipid for forming an empty nanodisc, and then the best ratio of lipid to TM protein to MSP for incorporating TM567. TM567 is a 23.6kDa protein with 206 residues, and its expression and purification has been described in chapter 3. All the experiments were carried out using one particular lipid POPC (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine) obtained from Anatrace, which is also a component in natural lipid bilayers.

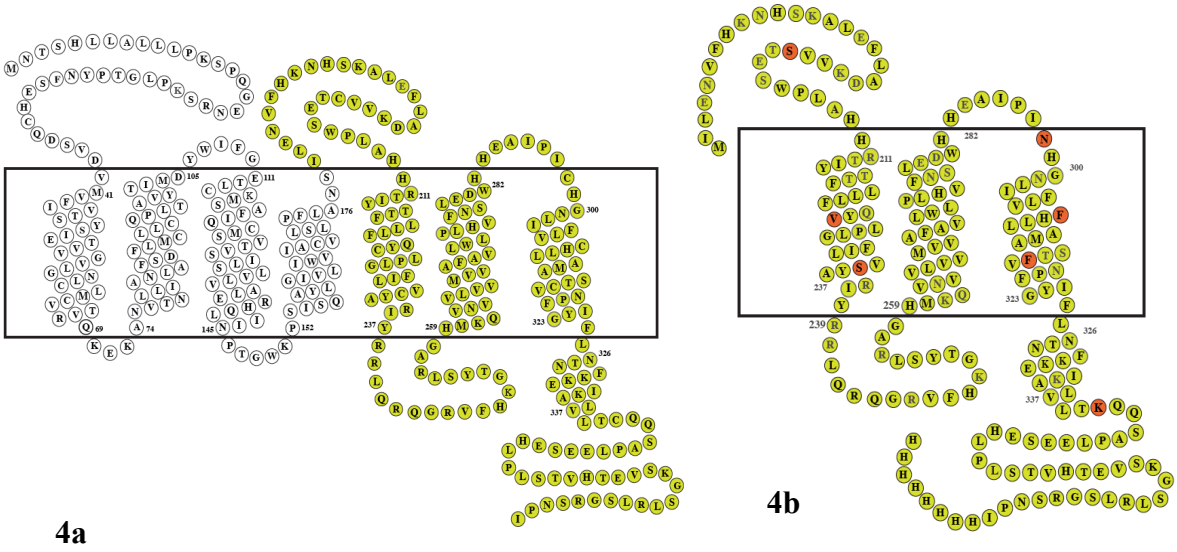


Figure 4. Snake plot of the Y4 receptor with the TM567His region highlighted in yellow (4a) and its Cys-free mutant (4b) illustrating the modified Cys-residues in red.

3. Results

Before setting-up experiments for incorporating the TMP into nanodiscs, we had studied the behavior of each constituent individually that is involved in the formation of nanodiscs.

3.1 Membrane Scaffold Protein (MSP): We have received MSP constructs of different sizes from the group of Gerhard Wagner at Harvard. Since we are interested to incorporate a 3-TM fragment of NPY4R, we have selected a smaller version of MSP with a deletion of helix 5. We name this construct MSPΔH5. This construct in vector harboring a kanamycin resistant gene was transformed into *E. coli* BL21 (DE3) tuner strain for expression in LB. Cells were induced with 1mM IPTG at a cell density (OD_{600}) of 0.7, allowing protein overexpression for 5 hrs at 37^{o}C. Subsequently, cells were pelleted, resuspended in Tris buffer (40mM Tris pH 8.0, 300mM NaCl), lysed using French press and the supernatant was subjected to affinity purification. All purification steps were carried out at 4^{o}C. Eluted fractions from affinity purification were pooled, subjected to dialysis to remove imidazole, and then TEV protease was added to remove the His-tag attached to MSPΔH5. After allowing for cleavage over 4 hrs, the reaction mix was again subjected to affinity purification and the purified MSPΔH5 without a His-tag was collected in the flow-through, which is analyzed by SDS-PAGE (Fig 5).

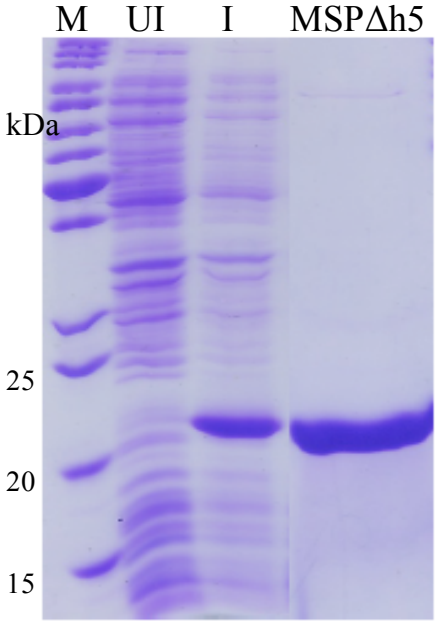


Figure 5. Purification of MSPΔH5 analyzed on SDS-PAGE. M-Marker, UI-Uninduced, I-Induced and the last lane is the purified MSPΔH5 after cleaving the His-tag with TEV protease.

3.2 TM567 fragment solubilized in detergents: The purification method was explained in the previous chapter. In brief, the purified and lyophilized TM567 is solubilized in an organic solvent (HFIP) and then mixed with the detergent (DPC) buffer. Around 3-5% DPC (100-150mM) was used to solubilize 1-2mg of TM567. The final volume is made up to 1 ml with water, subjected to sonication for 15 mins and then lyophilized. Lyophilized products are solubilized in plain water (200µl), subjected to sonication again and the behavior of the fragment was analyzed on a semi-preparative SEC pre-equilibrated in the same buffer containing DPC. To understand the effect of sonication, a similar experiment was set up without the sonication step, which led to the appearance of a monodisperse peak.

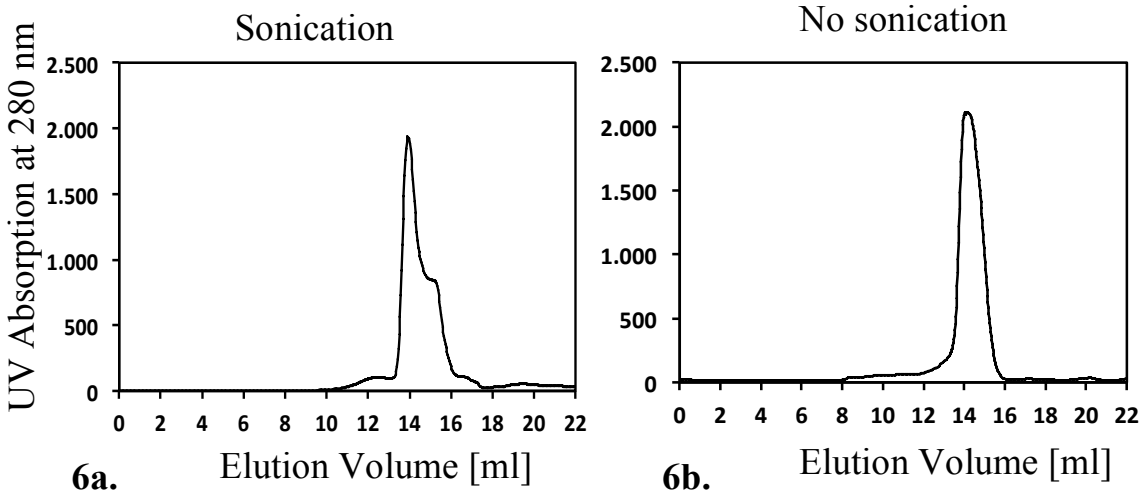


Figure 6. Behavior of TM567, analyzed by SEC, when solubilized in DPC detergent and its effect with (6a) and without sonication (6b).

From the SEC chromatograms we understood that the TM567 fragment exists in multiple states if sonicated but exists in monodisperse form if the sonication step is avoided. Thus for all subsequent experiments, we have eliminated the sonication step.

3.3 Detergents binding to Bio-beads observed by NMR: There are 2 types of detergents that have to be removed to allow for the formation of nanodiscs. The first detergent is sodium cholate, which is used for solubilizing lipids (POPC or DMPC) in water. The second one is DPC that was used for solubilizing the TMP. Although detergent removal can be done via dialysis, it is an extremely slow process and sometimes it requires more than a day for its complete removal. A good and efficient strategy is the use of Bio-beads, which trap the detergent molecules specifically and irreversibly, but not the lipid molecules. These beads are polystyrene-based available commercially as SM-2 (BioRad – used in our lab)[9,20] or Amberlite XAD2[21]. An experiment using NMR was done to observe the rate of detergent binding to the Bio-beads (fig 7). After incubating the detergent (100 mM DPC) for 3 hrs in presence of Bio-beads (0.3 mg/ml) at 4°C, 50% of the signals disappeared indicating that required the incubation time is around 5-6 hrs. Also tested was the quantity of Bio-beads required for absorbing a particular quantity of detergent. In the later experiments, we used 0.6 mg/ml of Bio-beads to reduce the incubation time to 2-3 hrs.

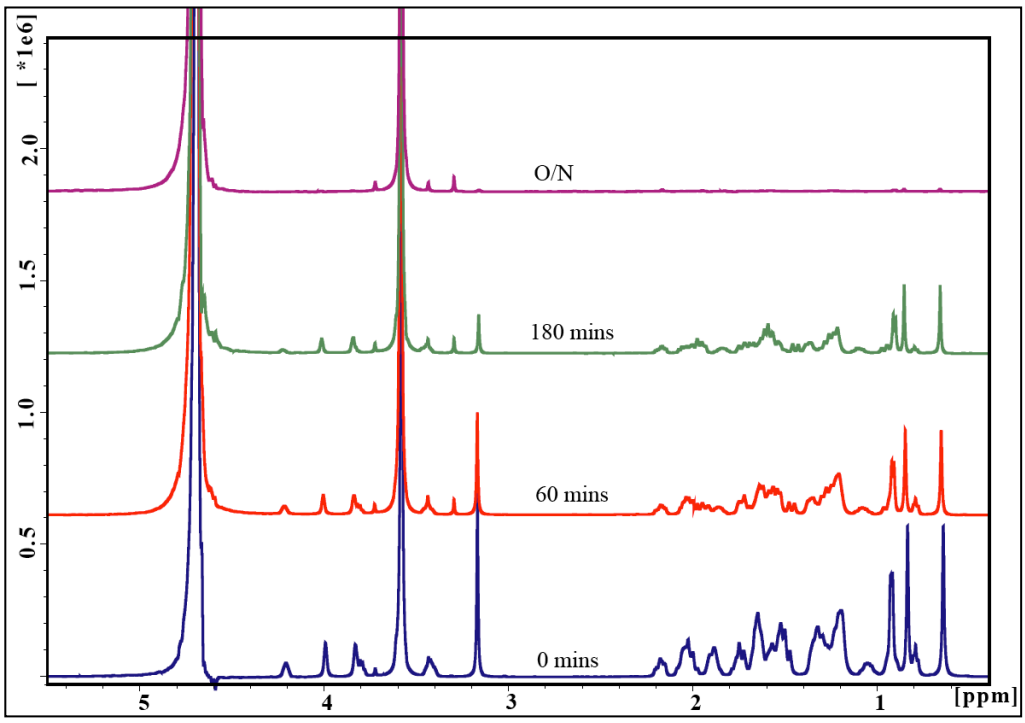


Figure 7. NMR spectra of the detergent DPC in presence of Bio-beads at different time intervals. Detergent was dissolved in Tris buffer and a ^{1}H NMR spectrum was recorded on 500 MHz spectrometer before (0') and after the addition of Bio-beads (60', 180' and

overnight incubation). Detergent peaks completely disappeared after incubating in Bio-beads for overnight.

3.4 Identifying the best MSP/lipid ratios for generating empty nanodiscs: The basic methodology involves solubilizing lipids (POPC) in presence of the detergent sodium cholate containing phosphate buffer (10 mM sod. phosphate pH 7.4 with 150 mM NaCl). This is followed by the addition of MSPΔH5, after which the mixture is gently mixed at 4^{o}C for 2 hrs. The starting concentrations used were 0.2 mM MSPΔH5, 10 mM lipids (MSP/Lipid ratio was 1:50). After gentle mixing for sufficient time (2 hrs.), optimized quantities (0.6 g/ml) of Bio-beads were added to remove the detergent under mild shaking for another 3-4 hrs. As the detergent binds to the Bio-beads, lipid bilayers are formed which later are enveloped with 2 molecules of MSP resulting in compact structures – the nanodiscs. After incubation, Bio-beads are allowed to settle at the bottom and the supernatant is filtered, concentrated and analyzed by SEC. The nature of the formed nanodiscs was analyzed using SEC. Based on the elution profiles, MSP/Lipid ratios are adjusted and followed by SEC. From various optimizations, we concluded that the MSP/Lipid ratio of 1:40 gave the best elution profile for the reconstitution of empty nanodiscs (fig 8).

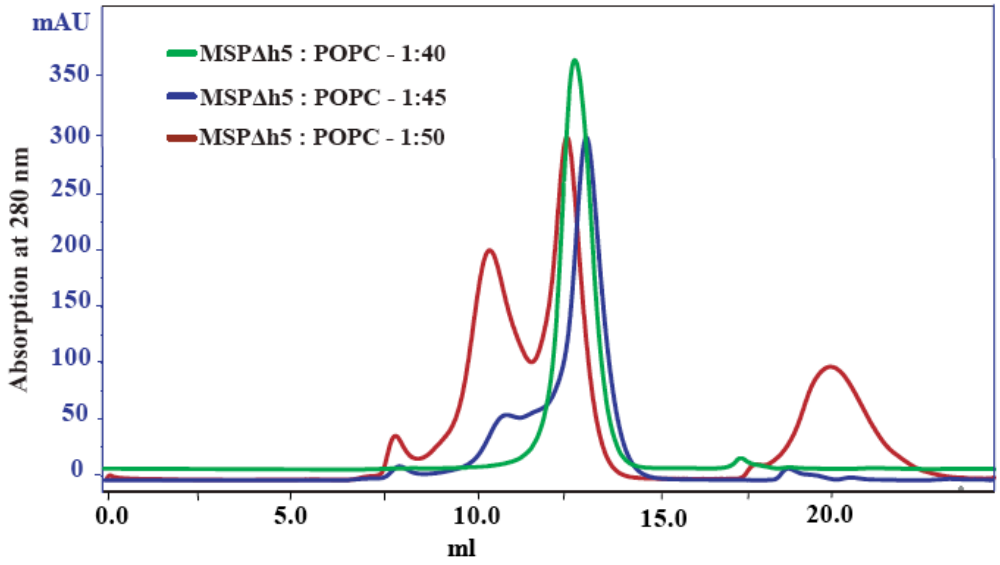


Figure 8: SEC profiles for the optimization of MSP/Lipid ratio. At a ratio of 1:50 (red) nanodiscs of different sizes/types occur, at a ratio of 1:45 (blue) larger sized nanodiscs disappear and at a ratio at 1:40 (green) monodisperse nanodiscs are found in the solution.

At this point it was also worthwhile for us to know how a pure MSP behaves on the SEC (fig 9), which would be useful for comparing the chromatograms obtained while optimizing reconstitution of nanodiscs. As expected, free MSP remains in aggregated form because of its amphipathic nature. This property has been observed when we loaded the purified MSPΔH5 on the SEC (fig 9). Free MSP runs as multimers on the SEC and we assume that the detergent present along with lipids can disrupt these multimers allowing for nanodisc reconstitution. Although different peaks are observed on SEC, they run as a single band on a reducing SDS-PAGE.

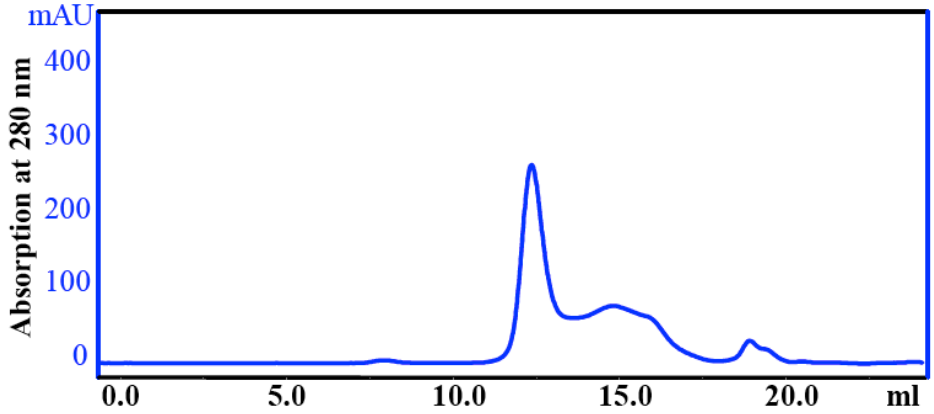


Figure 9. SEC profile of free MSPΔH5.

3.5 Reconstitution of the fragment TM567 into nanodiscs: The basic methodology is similar to the one followed for nanodisc reconstitution. Here we additionally mix lipids solubilized in cholate and MSP with TM567 solubilized in DPC. The concentrations involved were 0.2 mM TM567, 0.4 mM MSPΔH5 and 8 mM POPC lipids. The ratio of fragment/MSP/lipid was 1:2:40. We decreased the amount of lipids compared to amounts used for the reconstitution of the empty nanodiscs assuming that a small portion of lipids will be replaced by TM protein. After incubating the mixture for around 2 hrs at 4^{o}C, Bio-beads (0.6mg/ml) were added and again further incubated for 2-3 hrs under gentle shaking. This step removes both DPC and sodium cholate. Subsequently, Bio-beads were

removed by sedimentation, allowing purification of empty and TM-loaded nanodiscs. Nanodiscs formed were again analyzed by SEC (fig 10).

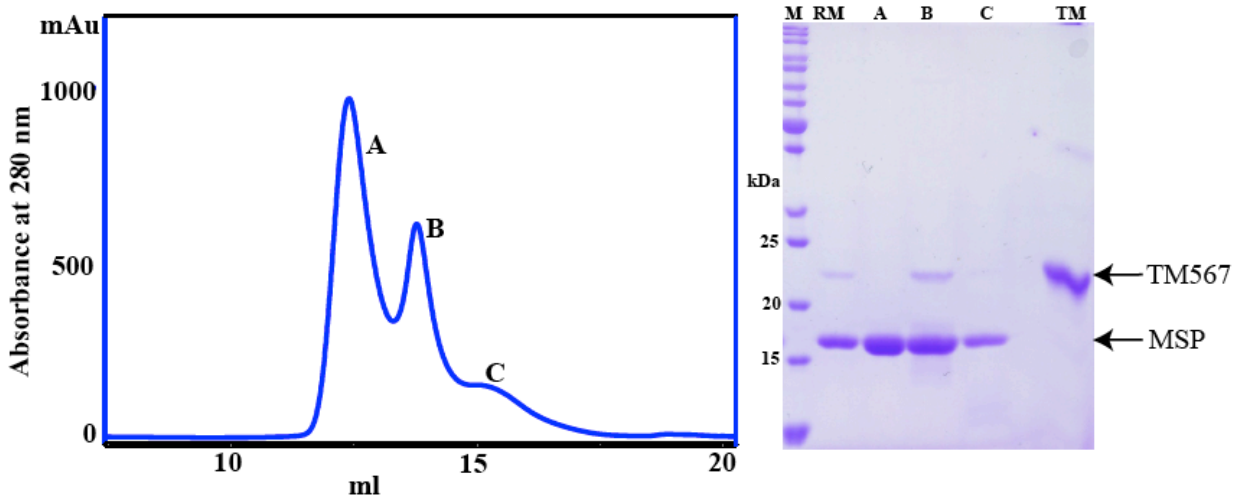
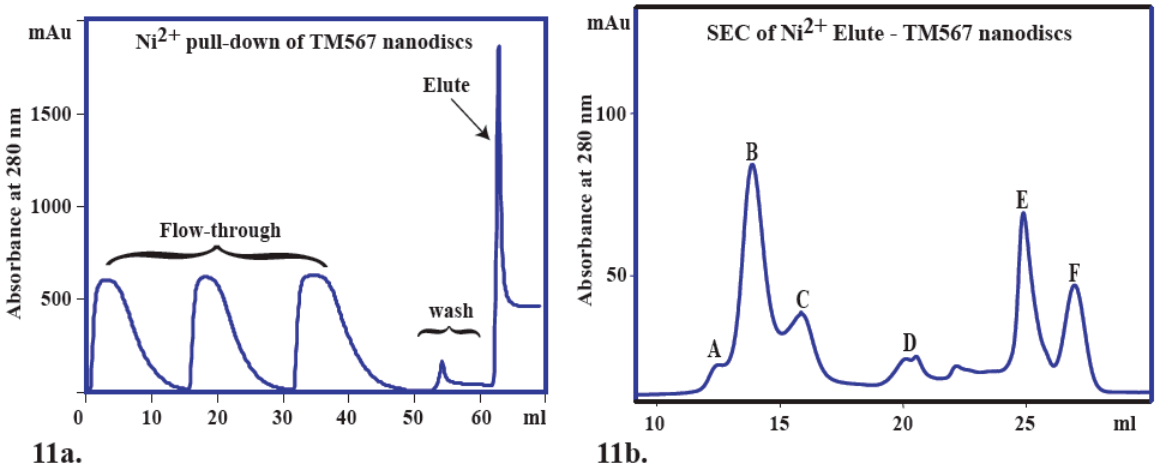


Figure 10. SEC profile of TM567 reconstitution into nanodiscs, and its analysis on SDS-PAGE. RM – reaction mix after removing Bio-beads; A, B and C are the peaks as observed on the SEC and TM = TM567 as a control.

Three peaks were observed on the SEC chromatogram and clearly the peak A was identical to the free MSPΔH5 as observed previously (fig 9), peak B shows the incorporated TM567 fragment as analyzed from SDS-PAGE and peak C, which is of low intensity, may be due to empty nanodiscs. This experiment indicates that we are capable of reconstituting TM567 into nanodiscs.

Since MSP is without His-tag and TM567 with His-tag, we decided to separate the loaded nanodiscs by Ni-NTA affinity purification. We followed the same procedure for the formation of the nanodiscs and, after removing the Bio-beads, the reaction mix was loaded onto a Ni^{2+} column. Empty nanodiscs are collected in the flow-through and the loaded nanodiscs are bound to the column, and eluted using 200 mM imidazole (fig 11a). The fractions from the affinity purification are analyzed on SDS-PAGE (fig 11c). Elute fractions from the affinity purification were pooled, concentrated and injected into a semi-preparative SEC (fig 11b).



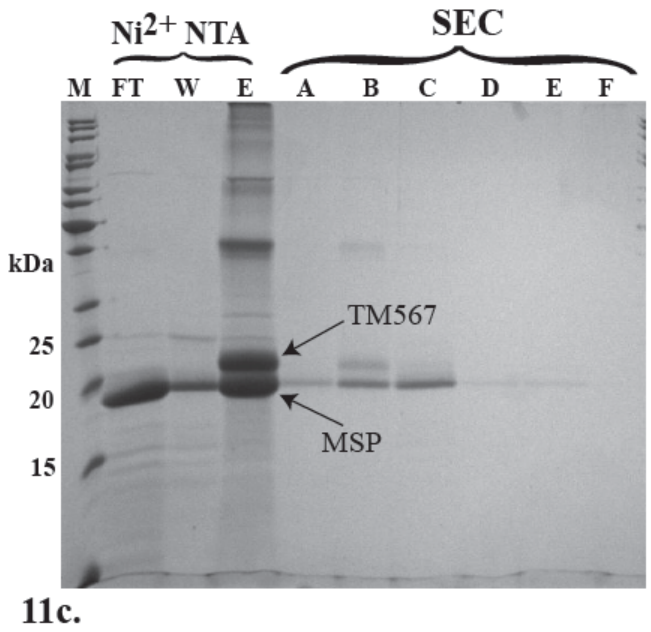


Figure 11. Ni^{2+}-NTA purification for separating the loaded nanodiscs (11a) and the Ni^{2+} elute containing the TM567 fragment from the SEC (11b). All peaks from the Ni^{2+}-NTA purification and SEC were analyzed using SDS-PAGE (11c).

The SEC purification confirmed the presence of loaded nanodiscs in peak B, which is the major peak, but peaks A and C could also contain loaded nanodiscs with different forms or shapes as we have loaded the Ni^{2+}-elute into the SEC. Peaks A and C are the minor peaks with less intensity and thus we could not see TM567 band on the SDS-PAGE. Although other peaks such as D, E and F are observed on the SEC, they didn't show any protein bands on the SDS-PAGE. These peaks could be related to the buffer components such as imidazole. We assume that peak B or peak C should contain the TM567

incorporated nanodiscs in monodisperse form. In our next experiments, we tried to change the MSP to lipid ratio to obtain Peak C as the major peak keeping the TM fragment concentration constant.

3.6 Identifying the right fragment/MSP/lipid ratios for obtaining monodisperse reconstitution of nanodiscs: In the previous experiment, we used the fragment/MSP/ lipid ratio as 1:2:40 and the major peak observed was B. To transfer the major peak to C, we assumed that it requires more quantities of MSP and lipids in the solution. Thus we increased the MSP content by 10 times and the lipid content by ~20 times to obtain a final ratio of 1:20:900. After removing Bio-beads, loaded nanodiscs were separated by affinity purification followed by SEC (fig 12a) and SDS-PAGE (fig 12b).

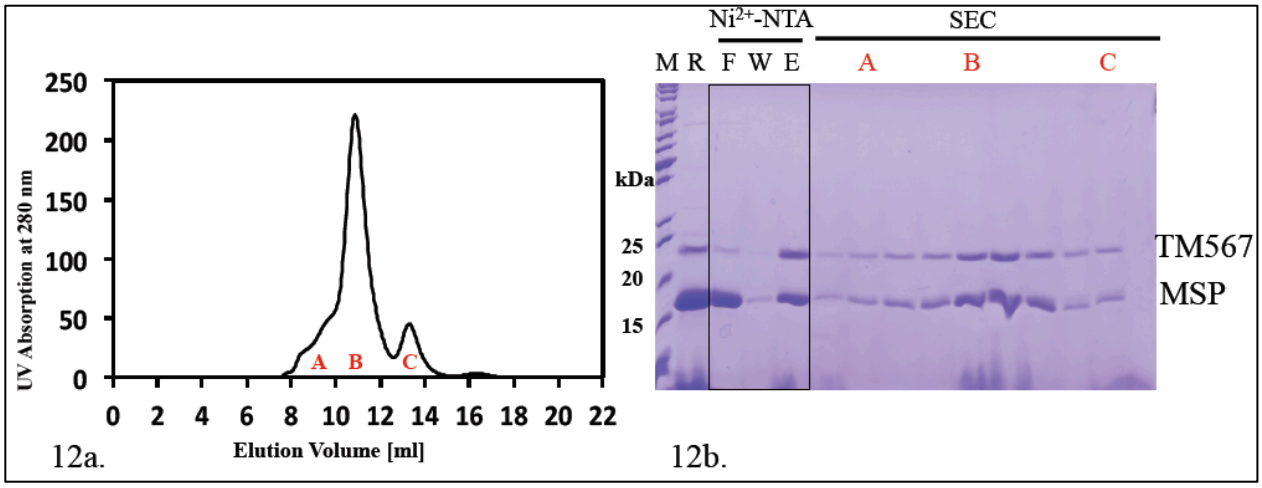


Figure 12. SEC profile of TM567 loaded nanodiscs obtained after affinity purification (12a). These nanodiscs are reconstituted at Fragment/MSP/Lipid ratios of 1:20:900. Observed peaks on SEC are analyzed using SDS-PAGE (12b). M-Marker; R-reaction mix obtained after removing Bio-beads; F- flow-through; W-wash; E-elute; (F, W, E – fractions from Ni^{2+} -NTA purification); A, B and C are all the corresponding fractions obtained from the SEC.

On increasing MSP and lipid by 10 and 20 times respectively, we observed a shift of peak A to peak B, but our goal was to increase peak C at the expense of peak B. Although this was not the desired effect we concluded from this that a decrease in lipid concentration would serve our purpose. Thus, in our next experiment we decreased the

lipid concentration, keeping the fragment and MSP concentration same. The new fragment/MSP/lipid ratio tested was 1:20:700 following the same experimental procedures discussed above (fig 13).

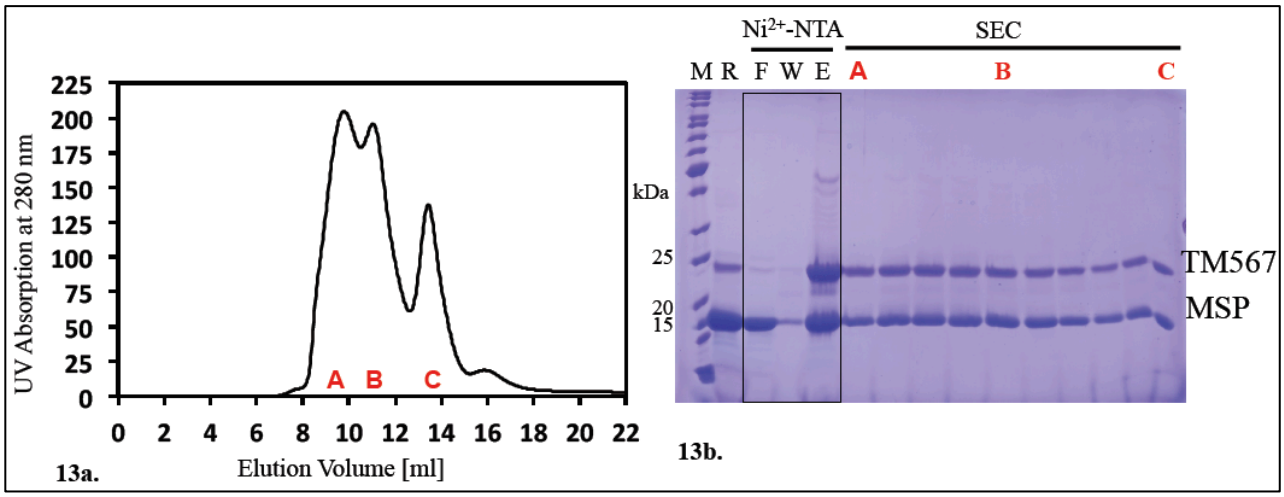


Figure 13. SEC profile of TM567-loaded nanodiscs obtained after affinity purification (13a). These nanodiscs are reconstituted at fragment/MSP/lipid ratios of 1:20:700. Observed peaks on SEC are analyzed using SDS-PAGE (13b). M-Marker; R-reaction mix obtained after removing Bio-beads; F- flow-through; W-wash; E-elute; (F, W, E – fractions from Ni^{2+} -NTA purification); A, B and C are all the corresponding fractions obtained from the SEC.

By decreasing the concentration of lipids, the equilibrium has shifted to peak A and thus peak A is the major peak. There is also an increase in the intensity of peak C, but it is still a minor peak. All 3 peaks A, B and C showed the reconstituted TM567 on SDS-PAGE analysis. From this experiment, it is clear that the change of lipid concentration should be in the opposite direction and thus in the following experiment MSP was increased 2 times and the lipid concentration was significantly increased compared to previous experiments. The new fragment/MSP/lipid ratios were 1:40:1600. Again the experimental procedure was similar to previous experiments. Fragment-incorporated nanodiscs were separated using affinity purification and their nature was examined on SEC (14a), followed by SDS-PAGE analysis (14b).

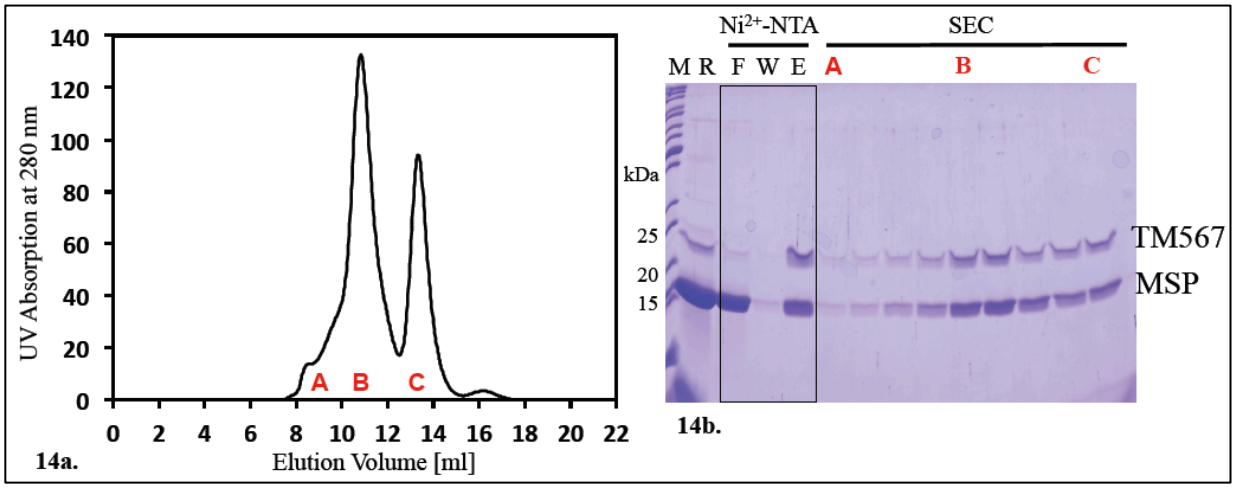


Figure 14. SEC profile of TM567 loaded nanodiscs obtained after affinity purification (14a). These nanodiscs are reconstituted at Fragment/MSP/Lipid ratios of 1:40:1600. Observed peaks on SEC are analyzed using SDS-PAGE (14b). M-Marker; R-reaction mix obtained after removing Bio-beads; F- flow-through; W-wash; E-elute; (F, W, E – fractions from Ni^{2+} -NTA purification); A, B and C are all the corresponding fractions obtained from the SEC.

With fragment/MSP/lipid ratios of 1:40:1600, the equilibrium definitely shifted to peak C, which was our desired direction, but still peak C is not the major component. One positive aspect is that peak A, likely due to large-sized, distorted nanodiscs, has completely disappeared. Probably the increase in peak C could be a shift in equilibrium from peak A. At this stage, peaks B and C are sufficiently separated allowing for separate biophysical analysis by NMR or electron microscopy.

Currently, scale-up of the previous experiment is under progress in our lab by collaborators.

3.7 Affect of Bio-beads on MSP and TMPs: In almost all experiments discussed above, we observed a decrease in protein concentration after the incubation with Bio-beads required for detergent removal. We speculated about binding of TM567 and MSP to the Bio-beads, and realized that other groups working on nanodiscs faced the same problem as well. Grisshamer and colleagues had also speculated the same about the adsorption of

neurotensin receptor (NTS1) to Bio-beads on prolonged incubation times or on excess usage of Bio-beads[22]. We monitored protein concentration by UV before and after the addition of nanodiscs. Protein concentration was measured at 280 nm before the addition of Bio-beads and was followed at regular intervals after Bio-bead addition (fig 15).

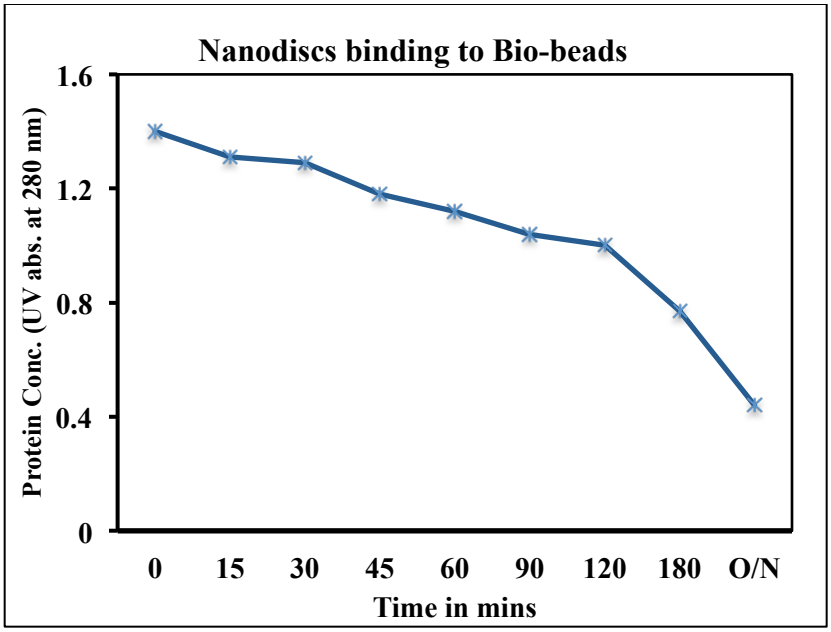


Figure 15. UV absorption profile of protein binding to Bio-beads. Protein concentration was measured at 280 nm in the nanodisc-forming reaction mixture that contains TM567 solubilized in DPC detergent, MSPΔH5, lipids dissolved in sodium cholate and Bio-beads.

It is very clear from the plot that the protein is absorbed by the Bio-beads. Concentration was reduced to more than half on incubating the reaction mix overnight. There was 20-30% loss of protein in 2 hrs, which is the normal incubation time followed after the addition of Bio-beads. The amount of added Bio-beads was chosen to match the rate at which they completely absorb the detergent in 2 hrs, after which they were immediately separated from the reaction mix to avoid further loss of the protein. It should also be noted that the decrease in protein concentration could be attributed to binding of TM567 or MSP independently to Bio-beads, which needs to be tested in future.

4. Discussion

We have described the procedure for incorporating TM567 of NPY4R into nanodiscs. Although more optimization is required to obtain the monodisperse peak as the majority, fragment/MSP/lipid ratios of 1:40:1600 gave the best results so far. From all the experiments mentioned above, it still difficult to predict the right combination of protein-to-lipid ratios in order to obtain monodisperse nanodiscs, or limit the formation of polydisperse liposome-like particles that can incorporate multimers of TMPs. After picking a particular combination, three common mandatory experiments are sequentially followed: affinity pull-down, size exclusion chromatography and SDS-PAGE analysis. On successful preparations using this protocol, biophysical analysis such dynamic light scattering can be done to reveal the hydrodynamic/Stokes radius pertaining to the size of the formed nanodiscs. These values could be very beneficial especially when setting-up NMR experiments for structural studies or for functional studies.

Incorporating TMPs into nanodiscs is a relatively simple process but to obtain homogenous monodisperse molecules, the process is highly determined by the type and combination of MSP and lipid used. The right combination for reconstituting one TMP may not be ideal for other TMPs. Many factors are involved in determining the right combination such as the size of the TMP or specifics of inter-helical or TMP-lipid interactions. The suitability in picking the right lipid or lipid combinations and most importantly the mode in which detergents used for solubilizing lipids are removed (dialysis or by Bio-beads) could definitely affect the folding and the nature of integration by TMPs into nanodiscs. Additional obvious factors that exist when working *in vitro* are pH, temperature and the choice of buffers, which contain the reducing agents. A slight variation in one of them could be critical and will change the equilibrium to irregularly shaped or aggregated or polydisperse nanodiscs (fig 16).

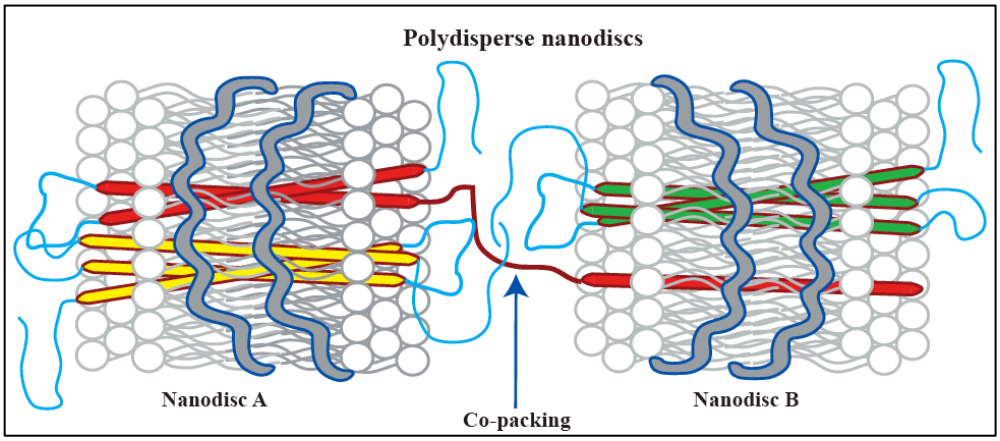


Figure 16. Schematic representation of aggregated or polydisperse nanodiscs. One of the helices of TMP (red) corresponding to nanodisc A is packed with the helices of TMP (green) in nanodisc B and in another case helices of one TMP (yellow) are packed against 2 helices of another TMP (red) within the same nanodisc resulting in complex form of an aggregated polydispersed nanodisc.

There could be many combinations that could be expected with regards to the packing of nanodiscs, but in our experiments we mainly see 3 forms on SEC corresponding to peaks A, B and C. As discussed above, we speculate that peak C corresponds to the monodisperse form of nanodiscs (fig 17a). Unfortunately, even after many optimizations this doesn't represent the majority of material. At a ratio of 1:40:1600 (fragment/MSP/lipid), peak C on the SEC profile constitutes about 40% of the total nanodiscs, and light scattering or NMR experiments are required to show their monodisperse nature. The major peak at ratios 1:2:40, 1:20:900 and at 1:40:1600 is always peak B on the SEC profile. Although the nature of the material corresponding to this peak is not entirely clear most likely it corresponds to nanodiscs incorporating 1-2 molecules of TM567. Alternatively, it may correspond to material in which helices from a single TMP are incorporated into different nanodiscs (fig 17b). Again this assumption has to be verified by experiment. At ratios of 1:20:700 where peak A was the main peak polydisperse nanodiscs packed in a complex manner are likely present (fig 17c). This peak was the major when the lipids were decreased from 900 mM to 700 mM while keeping TMP and MSP concentrations constant. This indicates that there is excess TMP and MSP with limited amounts of lipids required to form independent nanodiscs. This

results in recruiting more than one fragment in the nanodiscs and also co-sharing of helices between nanodiscs. Fragments existing in these nanodiscs are complicated to analyze, and the component ratios should be optimized to avoid forming these nanodiscs. At a given time, it's difficult to obtain 100% monodisperse species, as nanodisc formation is a self-assembly process, driven by all possible combinations (TMP, MSP and lipid) at different concentrations (unless perfectly defined for each component). Thus we have deduced a way to manipulate a crucial factor (MSP or lipid concentration) in obtaining nanodiscs of a distinct type.

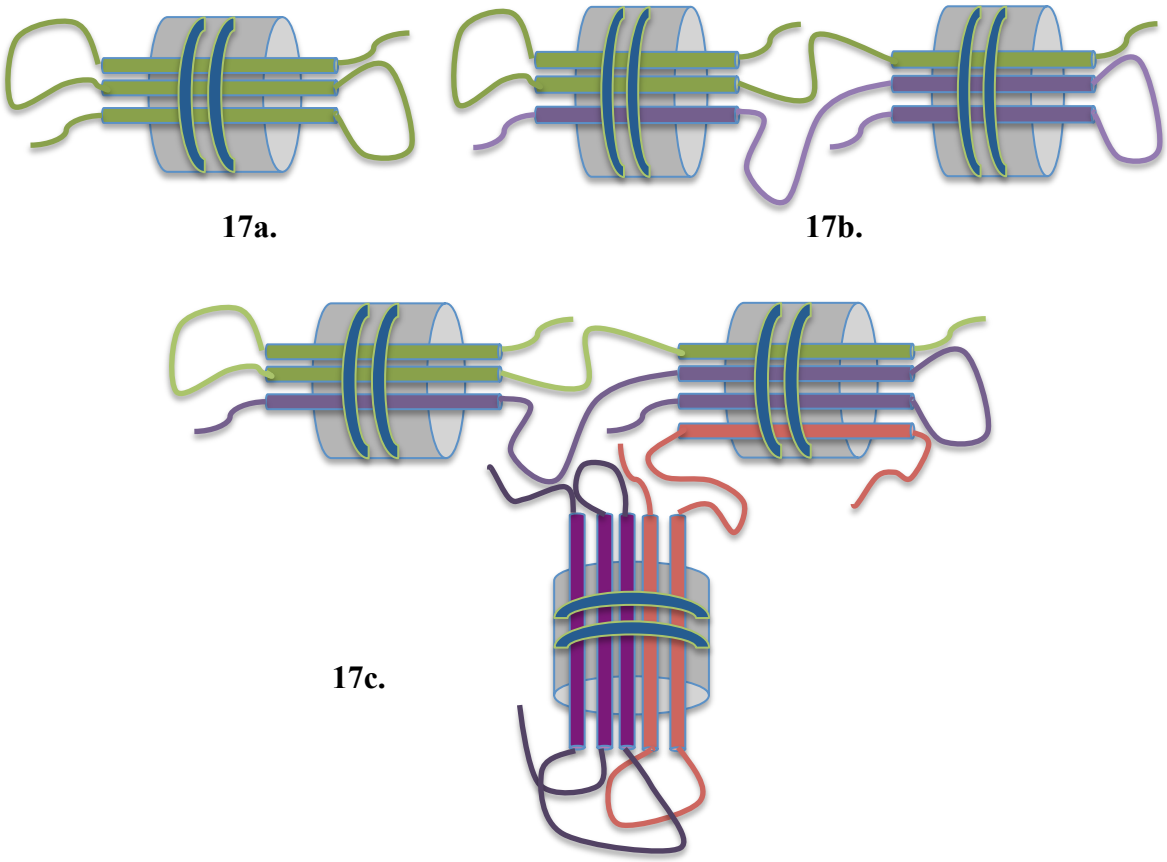


Figure 17. Different combinations of nanodiscs that can be formed depending on the MSP to lipid ratio. Obviously, monodisperse (17a) forms are desired for biophysical analysis.

Formed nanodiscs generally range from 9-15 nm in size and these could also be analyzed by electron microscopy, and many groups were previously successful in doing so. We

have also made a few attempts with empty nanodiscs but reasonable pictures were not obtained. Nevertheless we could use the expertise such as sample preparation, fixing and observing under EM for the incorporated nanodiscs. These electron micrographs can be of immense advantage before starting structural analysis by NMR or X-ray crystallography.

As discussed in the previous chapter, NMR signals for many transmembrane residues were missing. This could be attributed to the dynamic nature of micelles resulting in signal broadening. We hope that this can be overcome when the TM fragment is incorporated in nanodiscs, which are more rigid in nature. Hopefully we will be able to record NMR spectra of TM567 reconstituted into nanodiscs soon, which could be exciting to compare them to fragments incorporated in micelles.

5. References

1. Lee, A. G. Biological membranes: the importance of molecular detail. *Trends Biochem Sci* **36**, 493-500 (2011).
2. Opekarova, M. & Tanner, W. Specific lipid requirements of membrane proteins--a putative bottleneck in heterologous expression. *Biochim Biophys Acta* **1610**, 11-22 (2003).
3. Soubias, O. & Gawrisch, K. The role of the lipid matrix for structure and function of the GPCR rhodopsin. *Biochim Biophys Acta* **1818**, 234-240 (2012).
4. Spector, A. A. & Yorek, M. A. Membrane lipid composition and cellular function. *J Lipid Res* **26**, 1015-1035 (1985).
5. Smith, A. W. Lipid-protein interactions in biological membranes: a dynamic perspective. *Biochim Biophys Acta* **1818**, 172-177 (2012).
6. Serebryany, E., Zhu, G. A. & Yan, E. C. Artificial membrane-like environments for in vitro studies of purified G-protein coupled receptors. *Biochim Biophys Acta* **1818**, 225-233 (2012).
7. Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim Biophys Acta* **1666**, 105-117 (2004).
8. Ritchie, T. K., Kwon, H. & Atkins, W. M. Conformational analysis of human ATP-binding cassette transporter ABCB1 in lipid nanodiscs and inhibition by the antibodies MRK16 and UIC2. *J Biol Chem* **286**, 39489-39496 (2011).
9. Baas, B. J., Denisov, I. G. & Sligar, S. G. Homotropic cooperativity of monomeric cytochrome P450 3A4 in a nanoscale native bilayer environment. *Arch Biochem Biophys* **430**, 218-228 (2004).
10. Fang, Y., Gursky, O. & Atkinson, D. Lipid-binding studies of human apolipoprotein A-I and its terminally truncated mutants. *Biochemistry* **42**, 13260-

13268 (2003).
11. Krieger, M., Brown, M. S., Faust, J. R. & Goldstein, J. L. Replacement of endogenous cholesteryl esters of low density lipoprotein with exogenous cholesteryl linoleate. Reconstitution of a biologically active lipoprotein particle. *J Biol Chem* **253**, 4093-4101 (1978).
12. Denisov, I. G., Grinkova, Y. V., Lazarides, A. A. & Sligar, S. G. Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *J Am Chem Soc* **126**, 3477-3487 (2004).
13. Hagn, F., Etzkorn, M., Raschle, T. & Wagner, G. Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins. *J Am Chem Soc* **135**, 1919-1925 (2013).
14. Raschle, T. et al. Structural and functional characterization of the integral membrane protein VDAC-1 in lipid bilayer nanodiscs. *J Am Chem Soc* **131**, 17777-17779 (2009).
15. Yu, T. Y., Raschle, T., Hiller, S. & Wagner, G. Solution NMR spectroscopic characterization of human VDAC-2 in detergent micelles and lipid bilayer nanodiscs. *Biochim Biophys Acta* **1818**, 1562-1569 (2012).
16. Gluck, J. M. et al. Integral membrane proteins in nanodiscs can be studied by solution NMR spectroscopy. *J Am Chem Soc* **131**, 12060-12061 (2009).
17. Shenkarev, Z. O. et al. Lipid-protein nanodiscs as reference medium in detergent screening for high-resolution NMR studies of integral membrane proteins. *J Am Chem Soc* **132**, 5628-5629 (2010).
18. Bayburt, T. H., Leitz, A. J., Xie, G., Oprian, D. D. & Sligar, S. G. Transducin activation by nanoscale lipid bilayers containing one and two rhodopsins. *J Biol Chem* **282**, 14875-14881 (2007).
19. Bayburt, T. H., Grinkova, Y. V. & Sligar, S. G. Assembly of single bacteriorhodopsin trimers in bilayer nanodiscs. *Arch Biochem Biophys* **450**, 215-222 (2006).
20. Ritchie, T. K. et al. Chapter 11 - Reconstitution of membrane proteins in phospholipid bilayer nanodiscs. *Methods Enzymol* **464**, 211-231 (2009).
21. Nath, A., Grinkova, Y. V., Sligar, S. G. & Atkins, W. M. Ligand binding to cytochrome P450 3A4 in phospholipid bilayer nanodiscs: the effect of model membranes. *J Biol Chem* **282**, 28309-28320 (2007).
22. Inagaki, S., Ghirlando, R. & Grisshammer, R. Biophysical characterization of membrane proteins in nanodiscs. *Methods* **59**, 287-300 (2013).

Curriculum Vitae

Surname: KOCHERLA
Given name: Harsha Vardhan
Phone: +41-788829721;
Email: harsha.kocherla@chem.uzh.ch

Date of Birth: 21 July 1982
Nationality: Indian

CV Summary

- Employed as a PhD student since **June 2009** at the **Institute of Organic Chemistry**, University of Zurich, under the supervision of **Prof. Dr. Oliver Zerbe**.
- Research work involves biophysical characterization of membrane proteins (GPCRs) by NMR spectroscopy.
- Obtained my Masters degree from **Indian Institute of Technology – Madras.**
- Around 1-year professional experience in the Drug Safety department at Novartis.

PhD student – UZH, Switzerland **Jun 2009 – Present**

Thesis dissertation: Biochemical and Biophysical Characterization of Neuropeptide Y4 Receptor (a Human GPCR) by NMR. Important practical skills gained are:

- Technical expertise in measuring and analyzing NMR spectra (Proteins - 2D & 3D) independently to understand folding, dynamics, etc. both in micellar & lipidic environment.
- Expert at handling tough and labeled (^{15}N, ^{13}C, ^{2}H & GFP tagged) proteins such as protein complexes and membrane proteins (GPCRs) – both in denaturing system (Esp. using reverse phase HPLC) and membrane environment.

Publication: H. Kocherla, J. Marino, X. Shao, J. Graf, C. Zou, O. Zerbe: Biosynthesis and spectroscopic characterization of 2-TM fragments encompassing the sequence of a human GPCR, the Y4 receptor, ***ChemBioChem***, 2012 Apr 16;13(6):818-28

Professional Experience – Novartis, India **Apr 2008 – May 2009**

Description: Worked as a clinical safety-processing expert/ Argus Processing Expert in the Drug Safety and Epidemiology (DS&E) department. Nature of the work includes submitting Adverse Events (AEs) and Periodic Safety Update Reports (PSURs) of Novartis drugs already in the market to the global health authorities. Acted as ISEC officer and in implementing LEAN methodologies.

Project Assistant – Indian Institute of Technology, India **Aug 2007 – Feb 2008**

Guides – Dr. N.Manoj and Dr. G. Sathyanarayana Naidu

Project titled "*Biochemical and structural characterization of alr1529 lipase*".

Description: Alr1529 belongs to SGNH hydrolase superfamily, which is a class of α/β hydrolases. I was successful in generating three site directed mutants, study of enzyme deactivation kinetics and finally in setting up crystallization trials.

Masters – Indian Institute of Technology - Madras, India **Jul 2005 – Jun 2007**

Guide – Prof. Dr. A. Gopala Krishna

Thesis titled *"Molecular approaches to cardiac diseases: Functional aspects of a G - Protein Coupled Receptor and search for a natural product and its target molecule"*.

Bachelors – Osmania University, India **Jul 2005 – Jun 2007**

Major - Chemistry**; Minor –** Genetics and Microbiology

IT Skills: Microsoft Desktop Applications (Word, Power Point, Excel etc); NMR analysis tools – Topspin, CARA.

Extracurricular activities: Member of Badminton Club (ASVZ), organizing retreats and symposiums and participant of social, cultural, community and charity programs.

References

Prof. Dr. Oliver Zerbe
Institute of Chemistry
University of Zurich
Winterthurerstrasse 190,
8057, Zurich
Tel.: +41-44-635 42 63
oliver.zerbe@chem.uzh.ch

Dr. A. Gopala Krishna
Dept. of Biotechnology,
IIT Madras, Chennai – 600036
India
Tel: +91-44-2257-4112
agk@iitm.ac.in